DOCUMENT RESUME

ED 429 997 TM 029 707

AUTHOR	Zhu, Daming; Fan, Meichu
TITLE	Adjusting Computer Adaptive Test Starting Points To Conserve Item Pool.
PUB DATE	1999-04-00
NOTE	75p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE	Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Ability; *Adaptive Testing; *Computer Assisted Testing; Difficulty Level; *Item Banks; *Test Construction; Test Items

ABSTRACT

The convention for selecting starting points (that is, initial items) on a computerized adaptive test (CAT) is to choose as starting points items of medium difficulty for all examinees. Selecting a starting point based on prior information about an individual's ability was first suggested many years ago, but has been believed unimportant provided that the CAT is reasonably long. However, starting with a medium difficulty item for all examinees has two potential disadvantages: unnecessary uses of the first one or two items and overuse or overexposure of the items around the medium difficulty. This study analyzed simulated CAT results and suggests significant benefits from administering the first CAT item at a difficulty level suitable to each examinee. Such an adjustment can reduce the use of items around the medium difficulty in the item pool, providing extra help in controlling the exposure rate of the items beyond what standard exposure control methods can achieve. The effect of selecting examinee-appropriate starting points can vary depending on the quality of the information used about examinees' ability levels and the test termination rules applied. (Contains 3 tables, 26 figures, and 8 references.) (Author)

Adjusting Computer Adaptive Test Starting Points to Conserve Item Pool

Daming Zhu and Meichu Fan

ACT, Inc.

Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada, April 1999

Abstract

The convention for selecting starting points (that is, initial items) on a computerized adaptive test (CAT) is to choose as starting points items of medium difficulty for all examinees. Selecting a starting point based on prior information about an individual's ability was first suggested many years ago but has been believed unimportant provided that the CAT is reasonably long.

However, starting with a medium difficulty item for all examinees has two potential disadvantages: unnecessary uses of the first one or two items and overuse or overexposure of the items around the medium difficulty. This study analyzes simulated CAT results and suggests significant benefits from administering the first CAT item at a difficulty level suitable to each examinee. Such as adjustment can reduce the use of items around the medium difficulty in the item pool, providing extra help in controlling the exposure rate of the items beyond what standard exposure control methods can achieve. The effect of selecting examinee-appropriate starting points can vary depending on the quality of the information used about examinees' ability levels and the test termination rules applied.

Adjusting Computer Adaptive Test Starting Points to Conserve Item Pool

Introduction

Because of item response theory (IRT) and the general availability of computers, it has become possible to tailor a test by selecting questions of appropriate difficulty for each examinee. More and more research work in the educational measurement area has been focused on the promise and the problems relating to the computerized adaptive test (CAT) since the early 1970's. This is especially the case in recent years due to significant progress in computer technology and applications.

A CAT has many advantages. Among them are shorter tests (likely in both length and time) without loss of measurement accuracy, fewer motivational problems caused by questions of inappropriate difficulty, more convenient administration schedules, quicker reporting of test results, and new item types that would be difficult or impossible to do in paper-and-pencil tests (PPTs). However, there are also many challenges in planning, constructing, and administering a CAT. Test security, content validity represented by the items selected for each individual examinee, and measurement precision are among the measurement issues to be dealt with, in addition to facility and hardware concerns. Appropriate item selection, item exposure control, and item usage balance in an item pool, all of which have to do with test security, content validity, and measurement precision, are increasingly drawing researchers' attention.

Starting Points on a CAT

A CAT seeks to present items that are appropriate for each test taker in regard to the person's estimated level of skill or ability (Green, Bock, Humphreys, Linn, & Reckase, 1984). The convention for selecting a first item (or initial item) on a CAT is to choose an item of

medium difficulty (as a starting point) for each examinee when no information about the examinee's ability is known (Green et al., 1984; Hambleton, Zaal, & Pieters, 1991; Hulin, Drasgow, & Parsons, 1983; Wainer, 1990). The way the algorithm works is similar to the binary sort algorithm. Based on the examinees' performance on the initial item (whether the answer is correct or wrong), the ability estimate for the examinee is adjusted and the next item is selected based on the updated ability estimate. The same process continues in the selection of subsequent test items until the information collected regarding the examinee's ability reaches the established requirement or criteria for accuracy, at which point the test is terminated.

Hulin et al. (1983) discussed the options for selecting a starting point in CAT situations. They discussed two different approaches. In a relatively homogeneous examinee population and with little prior information about individual examinees' ability, it is reasonable to administer an initial item of moderate difficulty. When the examinee population is very heterogeneous, and information such as educational level can be obtained for the examinees before the test, an item of moderate difficulty appropriate for examinees with that particular educational level can be administered as the starting item.

Wainer (1990) further examined the starting point issue. He suggested using adjusted starting points for a certain group of examinees based on the information collected from groups of previous examinees with similar characteristics. He believed that a better guess of an examinee's ability could be made if more about that examinee is known—age, courses taken, and so forth. The information could be used to establish the initial estimate of proficiency the mean of some more narrowly defined group of previous examinees. A strategy exploiting auxiliary

information about examinees in this manner is better, in the sense of providing higher expected precision over the population of examinees.

Hambleton et al. (1991) stated that a good starting point would probably be one that is matched to the examinee's ability level. They suggested that information about the examinees' ability level, such as what can be inferred from educational background data or self-reports, could be helpful in deciding the starting point for each examinee. However, Hambleton et al. acknowledged that many researchers do not consider such adjustments necessary.

Lord indicated in his work in 1977, as reported by Hulin et al. (1983), that the choice of the starting item is relatively unimportant provided that the CAT is reasonably long--that is, has a variable length or fixed length with at least 25 items. The reasoning here is that the deviation of an inappropriate starting point in a CAT from the true ability will be narrowed down to a minimum and that the final measurement accuracy will not be compromised so long as there are enough items on the test.

Wainer and Kiely (1987) felt, however, that test anxiety and frustration are increased with inappropriate starting points. In addition, questions that are too easy or too difficult for the examinee contribute very little information about the person's ability (Green et al., 1984).

The Problem

In a population of which the abilities are normally distributed, a large number of examinees have their abilities around the medium level. Thus, in a CAT item pool, the usage and exposure of items with difficulties around the medium level could be very high. The convention of starting a CAT for every examinee by administering the first item at about medium difficulty has two potential disadvantages: unnecessary uses of the first one or two

items to various extents and overuse or overexposure of the items around the medium difficulty level. This puts high pressures on test developers to supply enough items around the medium difficulty level both for the initial item pool and for the later update and replacement of the item pool. In other words, starting with an item of average difficulty for all examinees could waste resources. If other information available about an examinee's ability level could be used for adjusting the starting point for the test taker, the starting items administered would be at a more appropriate difficulty level, and thus the use and exposure of items around the medium difficulty level would be reduced.

Purpose

The purpose of this study was to examine the impact on item usage of employing related information about examinees' educational background, such as courses taken and the course grades, to estimate each examinee's ability level and adjust the CAT starting point accordingly.

Method

The data used in the study were obtained from operational administrations of a large-scale standardized mathematics test. The data were from the administrations of nine different forms of the test, each of which contained six content areas and sixty discrete multiple-choice items in total. The whole data set contained approximately 30,000 examinees. Information on high school mathematics courses taken and grades earned by the examinees were collected (self-reported by examinees) when examinees registered for the test. Examinees' responses to

the test questions were scored. IRT parameters were estimated for each item and were calibrated across the nine test forms using BILOG (Mislevy & Bock, 1983).

Two mathematics educational background indices were computed based on examinee self-reported mathematics courses taken in high school and the corresponding grades earned. The first index is the grade point average (GPA) over all mathematics courses taken. Possible courses taken include Algebra I (first-year algebra), Algebra II (second-year algebra), Geometry, Trigonometry, Calculus, and other math beyond Algebra II (excluding the courses already listed). The second index (Course&GPA) is the ability estimate index computed using a model established by regressing examinees' GPA for the first three courses listed above and the number of mathematics courses taken towards their performance on the mathematics test.

The examinees' abilities were estimated based on their performance on the mathematics test. The positions of each individual examinee's GPA and Course&GPA values on the corresponding distributions were converted to ability level estimates according to the examinee ability distribution. These ability level estimates were later used as the reference for selecting starting points on the CAT.

Computer runs were conducted to simulate the CAT processes for each subject. The Three-Parameter Logistic (3-PL) Model was used in the CAT simulations. Two thousand subjects were randomly selected from the data set. Two types of CAT administrations were simulated. In the first type of runs, two fixed-length CAT administrations were simulated; each had 15 items and 30 items, respectively. In the second type of runs, the CAT had variable test length, with a maximum of 45 items and a minimum of 10 items for each subject. The test could end either when a predetermined accuracy level was reached or when the maximum number of test items (45 items) were taken by an examinee. Two sets of variable-length CAT

simulations were conducted, with a minimum posterior variance (*Pv*-value) of 0.0625 (high precision, equivalent to r=0.97) as the stopping rule for one set and a *Pv*-value of 0.1500 (low precision, equivalent to r=0.92) for another set.

Several other factors were involved in the CAT simulation. First, item balancing rules ensured that for each of the subtests every examinee took the same proportion of items as is specified for the conventional PPT. Second, the Sympson and Hetter (1985) exposure control method was employed to control the item exposure rate. (In this approach, several thousand CAT administrations are simulated; following each simulation, the frequency with which each item was presented is tallied and compared to some subjective maximum exposure rate. The exposure parameters for items with frequencies of use exceeding the standard are then successively adjusted downward as the cycle of simulations continues. The cycle ends when the exposure parameters have stabilized and no items exceed the usage standard. The advantage of this approach is that it works well for items that discriminate well near the center of the ability distribution; however, the approach can fail to protect items that discriminate well in the tails of the distribution. In the simulation, we used 0.9 and 0.1 as our exposure rate.) Third, the ability estimates were updated following each item response. The succession of estimates obtained as the test proceeds are commonly termed as *provisional,* reflecting the fact that each estimate is based only on what is known about the examinee at that point in the process. Several methods for computing provisional estimates have been proposed, each with its own advantages and disadvantages. Maximum likelihood estimation (MLE) methods have the advantage of being relatively unbiased, at least when compared to Bayesian procedures (Lord, 1980). However, MLE estimates can not converge at perfect response or all incorrect response patterns. Bayesian estimates are always bounded,

but can be significantly biased. Taking into account the advantages and the disadvantages of the MLE and the bayesian methods, in our CAT mathematics test simulations we used a hybrid approach to estimation, employing the Bayesian method for provisional ability estimates, and the MLE method for the final ability estimate.

For each type and length combination of the CAT, simulations were conducted using three different methods. In the first run, the starting point of the CAT was around the medium level on the item difficulty distribution of all items in the pool. In the second run, the starting point was at the estimated ability level converted from the subject's GPA. In the third run, the starting point was at the ability level derived from each subject's Course&GPA index.

In each simulation, the items each examinee took were recorded. The frequency of use of each item was also recorded. The correlation coefficients were computed between the subjects' scores (θ) on the real mathematics test and their scores ($\hat{\theta}$) on the different simulated tests.

Results and Discussion

Distributions of Starting Points

Table 1 and Figures 1 and 2 show the characteristics and the distributions of the first items used under three different starting item methods (No-Info, GPA, and Course&GPA) and two exposure control settings (0.90 and 0.10). When 0.90 was the exposure control rate, the No-Info method used only two starting items for all subjects, with one item (a=1.7414, b= -0.0671, and c=0.1163) used 1781 times and the other (a=1.7072, b=-0.1983, and c=0.0903) used 219 times. Both the GPA and the Course&GPA method used 12 items, with item b's

ranging from -1.4892 to 1.9879; the highest single starting item usage was 475 times in the GPA method and 469 times in the Course&GPA method. At the exposure control rate of 0.10, the three methods (No-Info, GPA, and Course&GPA) used 5, 60, and 66 items, respectively, as starting items; the highest rates of usage for a single starting item were 997, 425, and 137, respectively.

(Insert Table 1 here)

(Insert Figures 1~2 here)

Obviously, when a starting point was selected without using any information regarding the subject's ability, as was the case using the No-Info method, an item with a medium difficulty (b value) and appropriate a and c values--the combination of which would likely provide the most amount information about the subject's ability--would be used. Thus, a limited number of items will be selected as starting items even with a more restricted exposure control. These items would be exposed to a very large number of examinees. When subjects' GPA or Course&GPA was referenced in the process of choosing starting points, the selection of starting items was spread to many more items, with difficulties corresponding to examinees' positions on the GPA or Course&GPA distribution. The exposure rates of the starting items were therefore greatly reduced. However, it must be noted, as can been seen in Figures 1 and 2, that an item with a high difficulty value (a=2.3166, b=1.9879, and c=0.1295) was very often used as the first item, particularly with the GPA method. We take this to be the result of many subjects reporting a GPA of 4.0, a result which might not be very accurate and reliable. In the Course&GPA method, the effect of many reported GPA's of 4.0 was likely offset by the

variable of the number of mathematics courses taken in the regression model, resulting in relatively lower usage of that particular high-difficulty item at the start. The inaccuracy in the GPA reported may come from two main sources: the incomparability of the grades across courses and schools, and the misreporting of grades by the examinees at the time they registered to take the PPTs.

Item Usage and Usage Distributions

Table 2 summarizes the results of the fixed-length (15-item and 30-item) test. The No-Info method used the least number of items in a test while the other two methods used about the same number of items. The differences were approximately 20 items between the No-Info and the other two methods when the exposure control rate was 0.90 and were about 12 items under the exposure control rate of 0.10. Consequently, the No-Info method had much higher mean item usage (the average usage over the items used) and maximum single item usage in simulations with an exposure control rate of 0.90. The mean of the item usage with the No-Info method was around 50 times more than that with the other two methods in the 15-item test and about 30 times more in the 30-item test. For the maximum individual item usage, the differences between the No-Info and the other two methods were approximately 600 times in a 15-item test and about 550 times in a 30-item test. When 0.10 was used for exposure control, the differences in these item usage statistics became closer between the No-Info method and the other two, with the latter two being about the same. In one situation, the 15-item test simulations, the No-Info method had a lower maximum item usage than did the other two methods.

(Insert Table 2 here)

Figures 3, 4, and 5 show the distributions of item usage in 15-item tests under three different starting point methods. It can easily be seen that the items with medium difficulty were used much more heavily in the No-Info method than they were in the other two methods. Between the GPA and Course&GPA methods, the distributions were very similar, with a few exceptions. The most noticeable exceptions were several heavy-usage points at the high difficulty end with the GPA method, which can be explained by the heavier influence of many reported GPA's of 4.0. The item distributions of item usage in 30-item tests are illustrated in Figures 6, 7, and 8, which show the same trend seen in the 15-item test results.

(Insert Figure 3 to 8 here)

In fixed-length (15-item and 30-item) CAT simulations with exposure control at the 0.10 level, the item usage distributions, as shown in Figure 9 through Figure 14, had different characteristics although the summary statistics from these simulations in Table 2 were not that much different. One difference was that the No-Info method tended to have more even item usage across the difficulty range, with somewhat heavier item usage in the middle one third of the item difficulty range of the items used. The other difference was the relatively higher single item usage found near one or both ends of the item difficulty range associated with the other two methods.

(Insert Figure 9 to 14 here)

In simulations for tests with variable length but a maximum of 45 items, the item usage distributions resembled those of the fixed-length tests. Table 3 and Figures 15 through 26 illustrate the item usage distributions in variable-length tests using different methods under different exposure control and test termination rule combinations.

(Insert Table 3 here)

(Insert Figure 15 to 26 here)

Less items were used in test simulations when all examinees started the test on items with medium difficulties, compared to the results when GPA and Course&GPA information was used in selecting starting points for examinees. This could result in overuse (overexposure) of some items in the pool, as indicated by the higher mean item usage and higher maximum single item usage associated with the No-Info method. This would likely happen in a CAT with weak exposure control measures. The differences among the three methods in the number of items used, mean item usage, and maximum single item usage among the methods would be reduced when stronger exposure controls were imposed, as shown in the simulations with exposure control rate of 0.10. However, the usage of the items in the medium difficulty range still tends to be heavier when the No-Info method was used, as is indicated in the illustrations of item usage distributions.

The higher single item usage of some items near the high and low ends of the item difficulty range associated with the GPA and Course&GPA methods likely came from the

particular distribution of the GPA and Course&GPA information. This result indicates that the quality of information used about each examinee's ability level would influence the appropriateness of the starting point decision and thus the effect of the CAT process.

Correlation of θ and $\hat{\theta}$

A comparison of the correlation of θ and $\hat{\theta}$ obtained from different starting methods (see Table 2) shows that those from the No-Info and the Course&GPA methods were close in most simulations. In the simulations, differences between the results of the two methods were tiny, with no clear patterns. The exceptions were found in the results of variable-length tests with the more relaxed termination rule (posterior variance = 0.15); where average test lengths were relatively short, the Course&GPA method produced ability estimates ($\hat{\theta}$) that correlate slightly higher to the true ability (θ) than the No-Info method did (0.914 vs. 0.907 under exposure control of 0.90; 0.913 vs. 0.903 under exposure control of 0.10). As to the average test lengths of the simulations, those using Course&GPA were a little longer (more than one item but less than two items on average) than those using the No-Info method were. These differences in correlation of θ and $\hat{\theta}$ may have come from the differences in average test lengths between the No-Info and the Course&GPA methods.

The scores associated with the GPA method had consistently the lowest correlation among the three methods in all simulations. The differences in the correlation of θ and $\hat{\theta}$ between the GPA method and the other two methods were as high as 0.06. These differences probably were caused by the inaccuracy of the GPA information, that is, an inconsistency between the examinees' GPA rankings and their true ability levels (θ), which was indicated by

the only moderate correlation coefficient (r=0.578) between the two. The correlation of θ and $\hat{\theta}$ obtained by using the GPA method was closest to those obtained by using the No-Info or Course&GPA methods in the simulations of 30-item tests.

The results confirmed the common understanding that for a CAT when a test length is long enough, the impact of inaccurate starting points diminishes. When GPA was combined with the number of mathematics courses taken, the quality of prediction information improved (r=0.695). Using Course&GPA information in selecting starting points on a CAT helped reduce the exposure of items around the medium difficulty levels, particularly in relatively short CATs, and achieved the same level of measurement accuracy, if not slightly better, compared to the results of using the No-Info method.

Conclusions

Efficiency and cost-effectiveness are among the technical and practical issues to be resolved before actual implementation of CATs. Using additional information about examinees' ability levels , when it is available, to select the first item on a CAT at a difficulty level suitable to each examinee can reduce the usage of items around the medium difficulty. This approach could provide extra help in controlling the exposure rate of the items in a CAT pool, beyond what standard exposure control methods do. The actual effect of selecting starting points can vary depending on the quality of the information about examinees' ability levels and on other factors, such as the exposure control and the test termination rules used. Further investigation in this area will certainly be necessary and shows promise for improving the accuracy and efficiency of CATs.

Reference

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement, 21*(4), 347-360.

Hambleton, R. K., Zaal, N. J., & Pieters, J. P. M. (1991). Computerized adaptive testing: theory, applications, and standards. In R. K. Hambleton & N. J. Zaal (Eds.), *Advances in Educational and Psychological Testing: theory and applications* (pp. 341-366). Boston: Kluwer Academic Publishers.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item response theory application to psychological measurement*. Homewood, IL: Dow Jones-Irwin.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Mislevy, R. J., & Bock, R. D. (1983). BILOG: Item and test scoring with binary logistic models [Computer program]. Mooresville, IN: Scientific Software.

Sympson, J.B., & Hetter, R.D. (1985). Controlling item-exposure rates in computerized adaptive testing. *Proceedings of the 27th annual meeting of the Millitary Testing Association* (973-977). San Diego, CA: Navy Personnel Research and Development Center.

Wainer, H. (1990). *Computerized adaptive testing: A primer*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24,* 185-201.

Table 1. Starting Item Statistics Summary

Exposure Control	Starting Item Method	# of Starting Items	Starting Item Difficulties (b)	Starting Item Usage
0.90	No-Inf	2	-0.198 ~ -0.067	219 ~ 1781
	GPA	12	-2.038 ~ 1.988	1 ~ 475
	Course&GPA	12	-2.038 ~ 1.988	14 ~ 469
0.10	No-Inf	29	-0.436 ~ 0.429	2 ~ 199
	GPA	60	-2.038 ~ 1.988	1 ~ 425
	Course&GPA	66	-2.038 ~ 1.988	1 ~ 137

Table 2. Summary Statistics for Fixed-Length Tests

Test Length	Exposure Control	Starting Item Method	$r(\theta, \text{1st-Item-}b)$	Item Used	Mean Usage	Usage SD	Max. Item Usage	$r(\theta, \hat{\theta})$
15-Item	0.90	No-Inf	-0.009	99	303	344	1781	0.944
		GPA	0.580	118	254	270	1144	0.918
		Course&GPA	0.686	120	250	267	1182	0.948
	0.10	No-Inf	0.028	216	139	81	240	0.934
		GPA	0.571	227	132	85	445	0.903
		Course&GPA	0.680	228	132	81	365	0.934
30-Item	0.90	No-Inf	0.040	181	332	331	1814	0.972
		GPA	0.580	200	300	298	1249	0.970
		Course&GPA	0.686	199	302	295	1269	0.973
	0.10	No-Inf	0.047	378	159	86	514	0.963
		GPA	0.576	391	154	86	468	0.954
		Course&GPA	0.678	391	154	86	509	0.960

Table 3. Summary Statistics for Variable-Length Tests

Test Stop Rule	Exposure Control	Starting Item Method	$r(\theta$, 1st-Item-$b)$	Item Used	Mean Usage	Usage SD	Max. Item Usage	Average Test Length	$r(\theta,\hat{\theta})$
Pv=0.0625	0.90	No-Inf	0.003	177	262	311	1803	23.2	0.964
(r=0.97)		GPA	0.580	215	217	256	1126	23.3	0.906
		Course&GPA	0.686	209	229	262	1241	23.9	0.963
	0.10	No-Inf	0.006	491	147	60	231	36.1	0.958
		GPA	0.566	498	138	60	435	34.3	0.947
		Course&GPA	0.672	492	145	59	249	35.6	0.961
Pv=0.1500	0.90	No-Inf	0.004	68	299	361	1805	10.2	0.907
(r=0.92)		GPA	0.580	127	174	228	1037	11.1	0.871
		Course&GPA	0.686	136	166	229	1186	11.3	0.914
	0.10	No-Inf	0.009	405	64	57	214	13.0	0.903
		GPA	0.572	438	65	55	425	14.2	0.877
		Course&GPA	0.675	452	64	52	191	14.6	0.913

Figure 1. Starting Item Usage Using Different Methods (Exposure Control at 0.90)

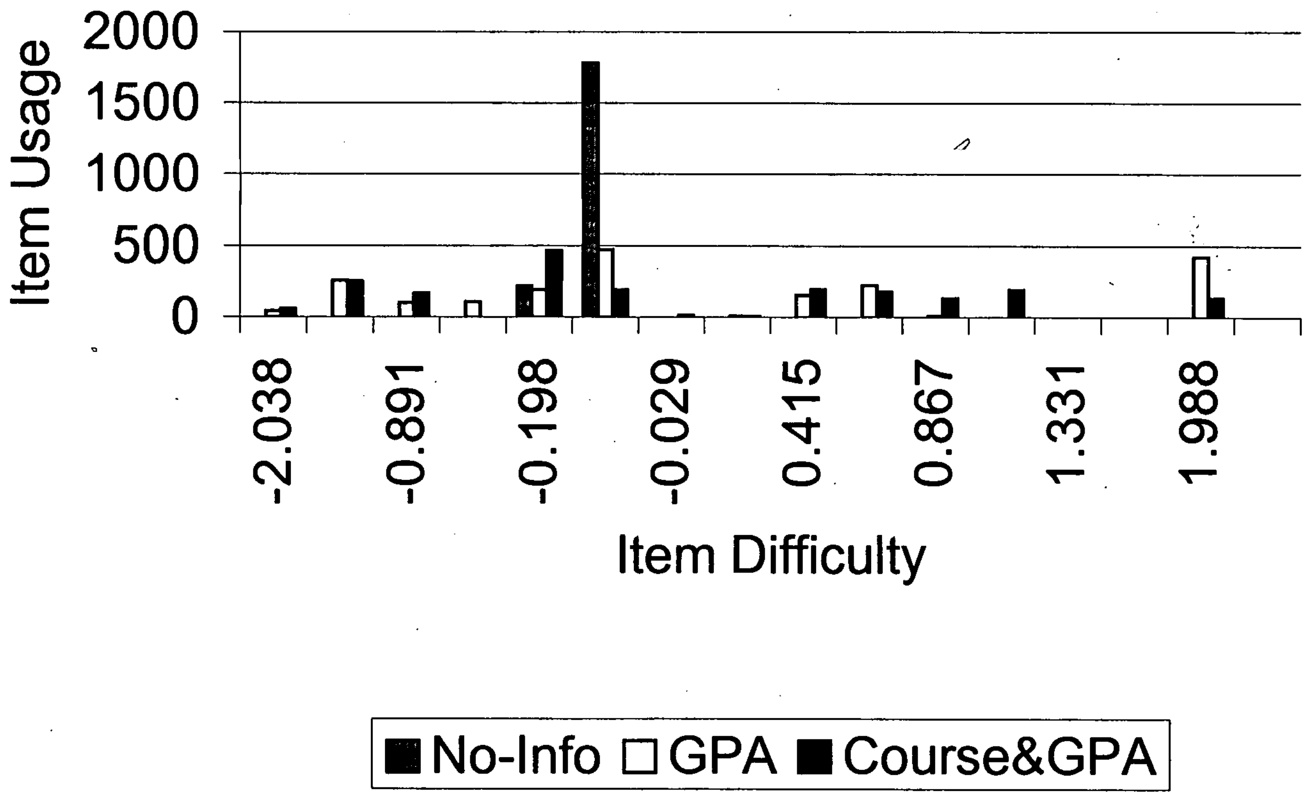


Figure 2. Starting Item Usage Using Different Methods
(Exposure Control at 0.10)

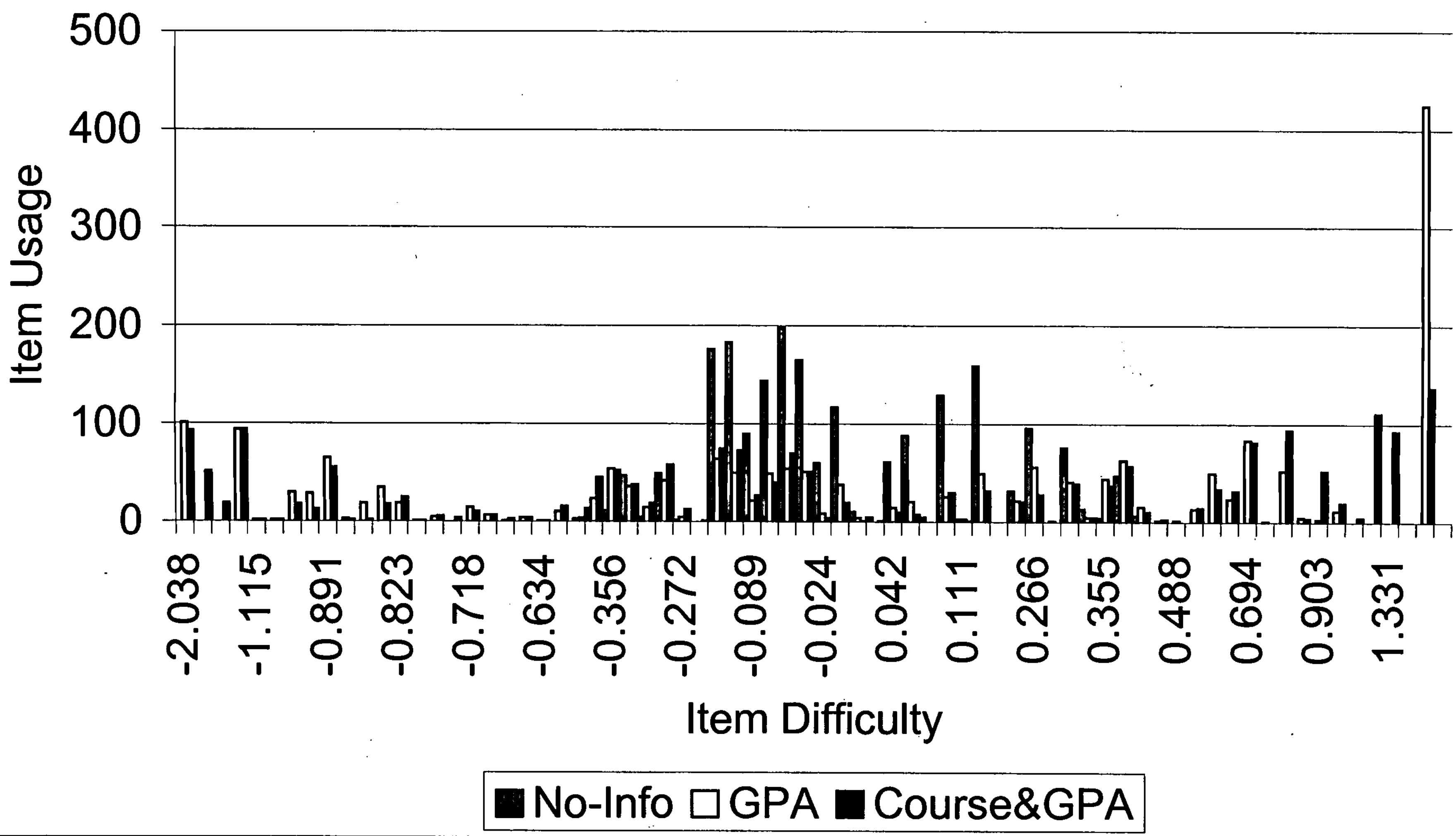


Figure 3. Item Usage Distribution on 15-Item Test (Exposure Control at 0.90)--No-Info Method

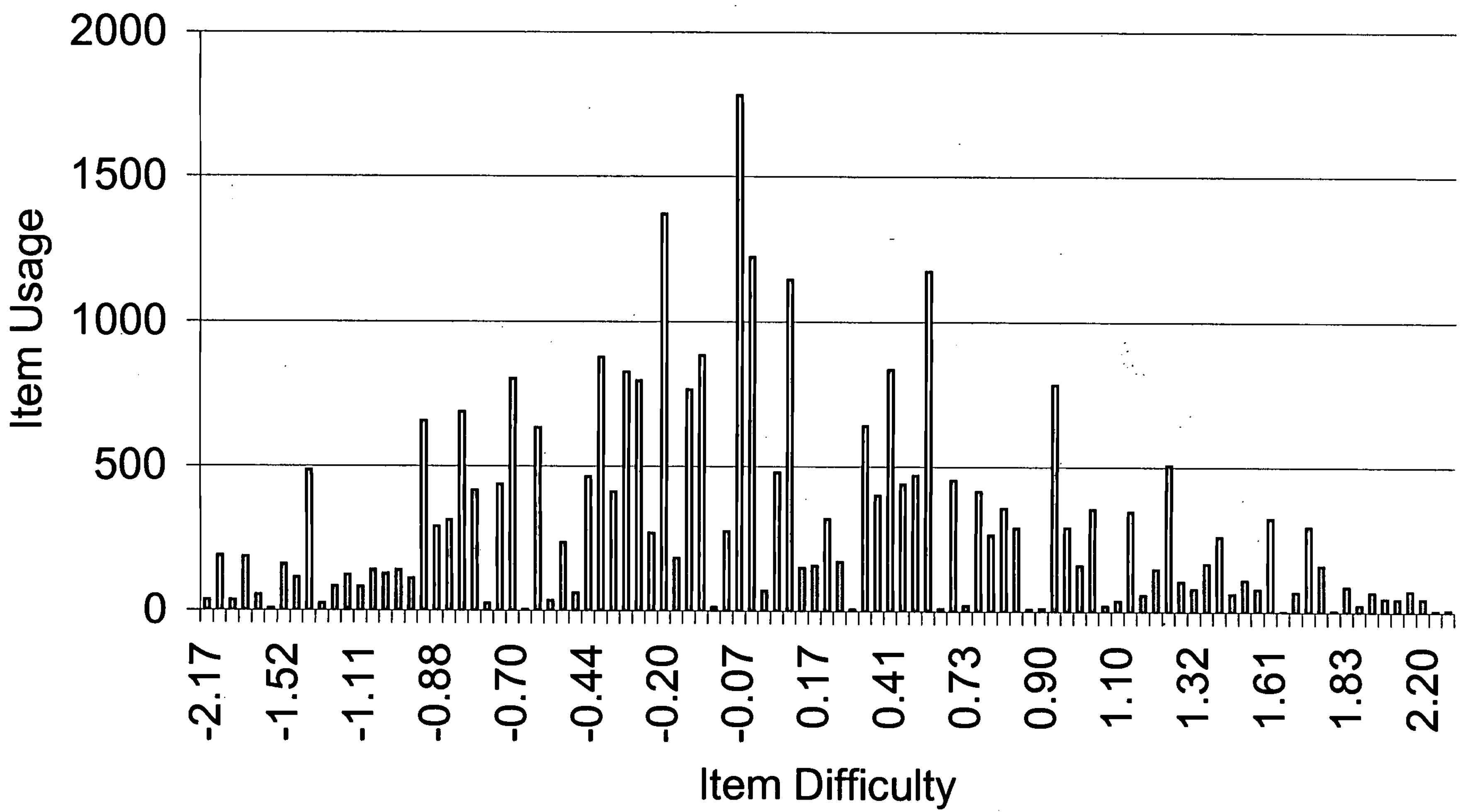


Figure 4. Item Usage Distribution on 15-Item Test (Exposure Control at 0.90)--GPA Method

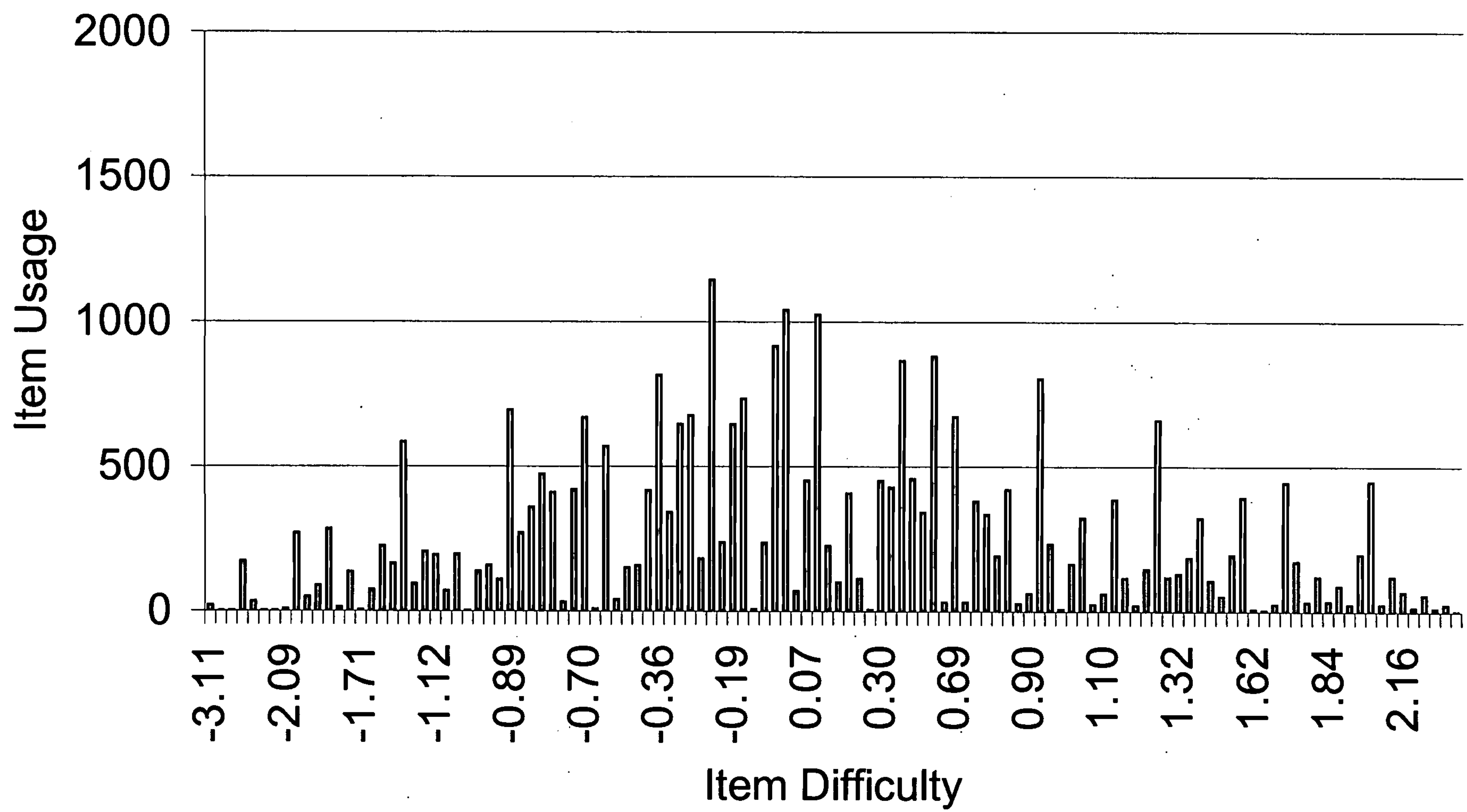


Figure 5. Item Usage Distribution on 15-Item Test (Exposure Control at 0.90)--Course&GPA Method

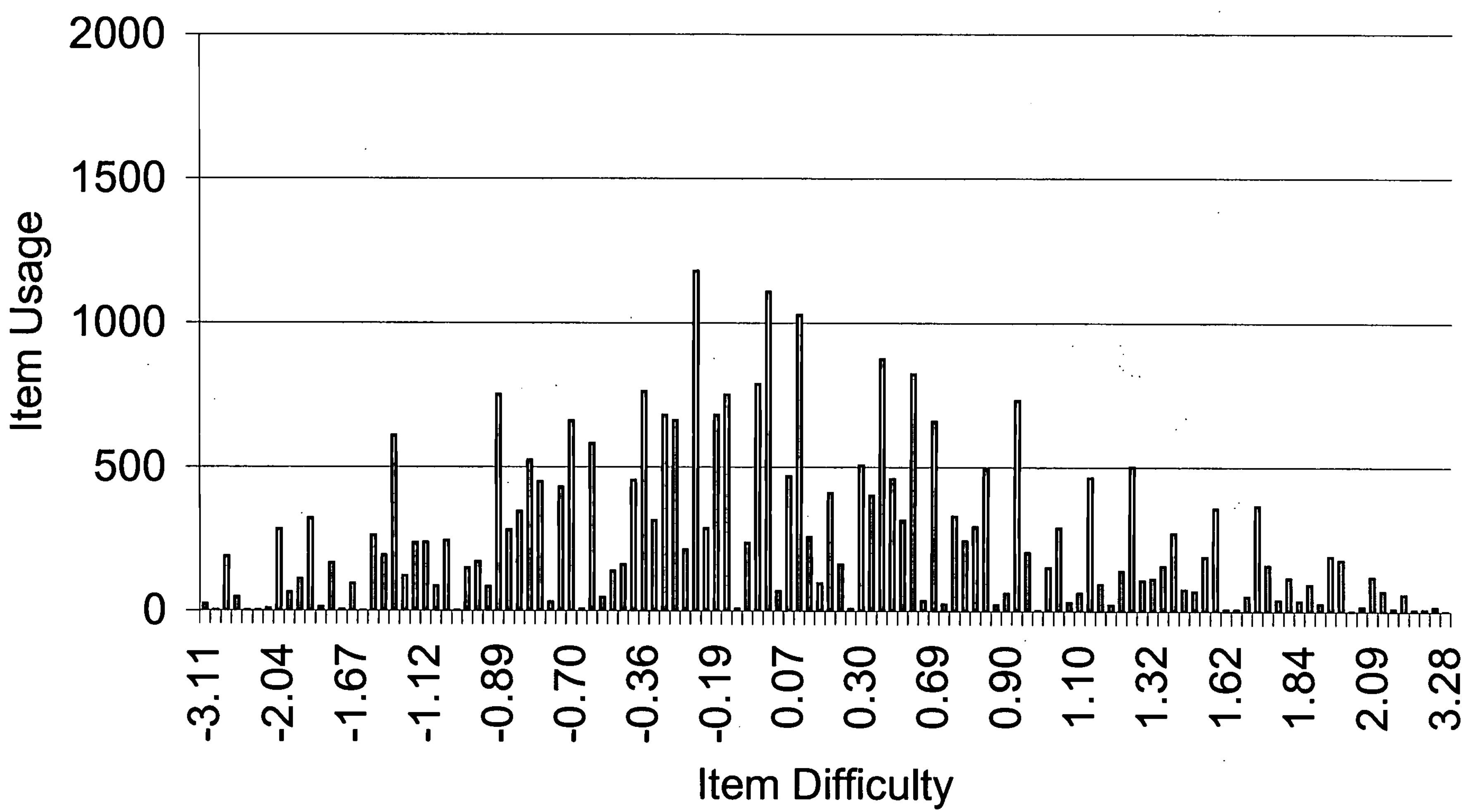


Figure 6. Item Usage Distribution on 30-Item Test (Exposure Control at 0.90)--No-Info Method

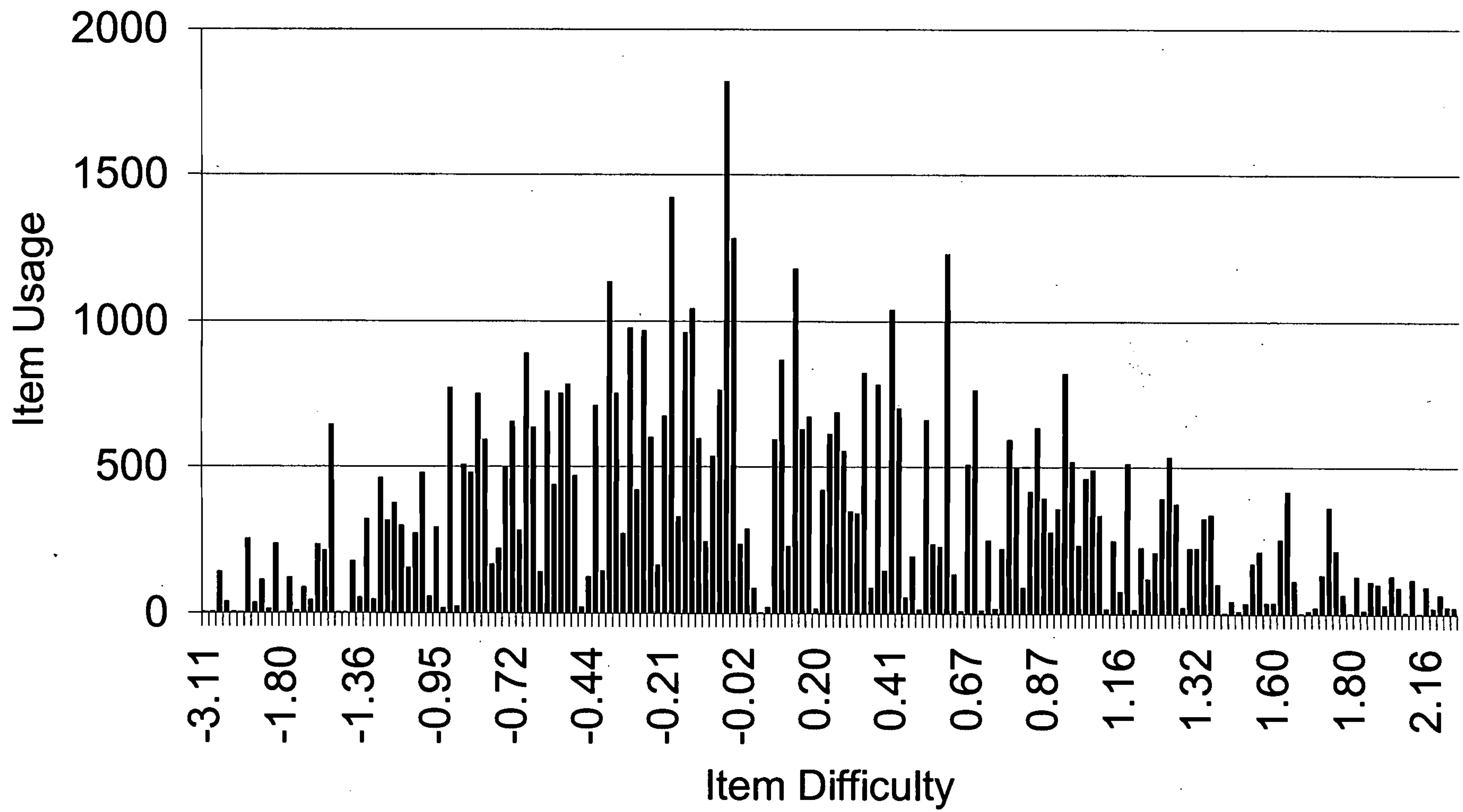


Figure 7. Item Usage Distribution on 30-Item Test (Exposure Control at 0.90)--GPA Method

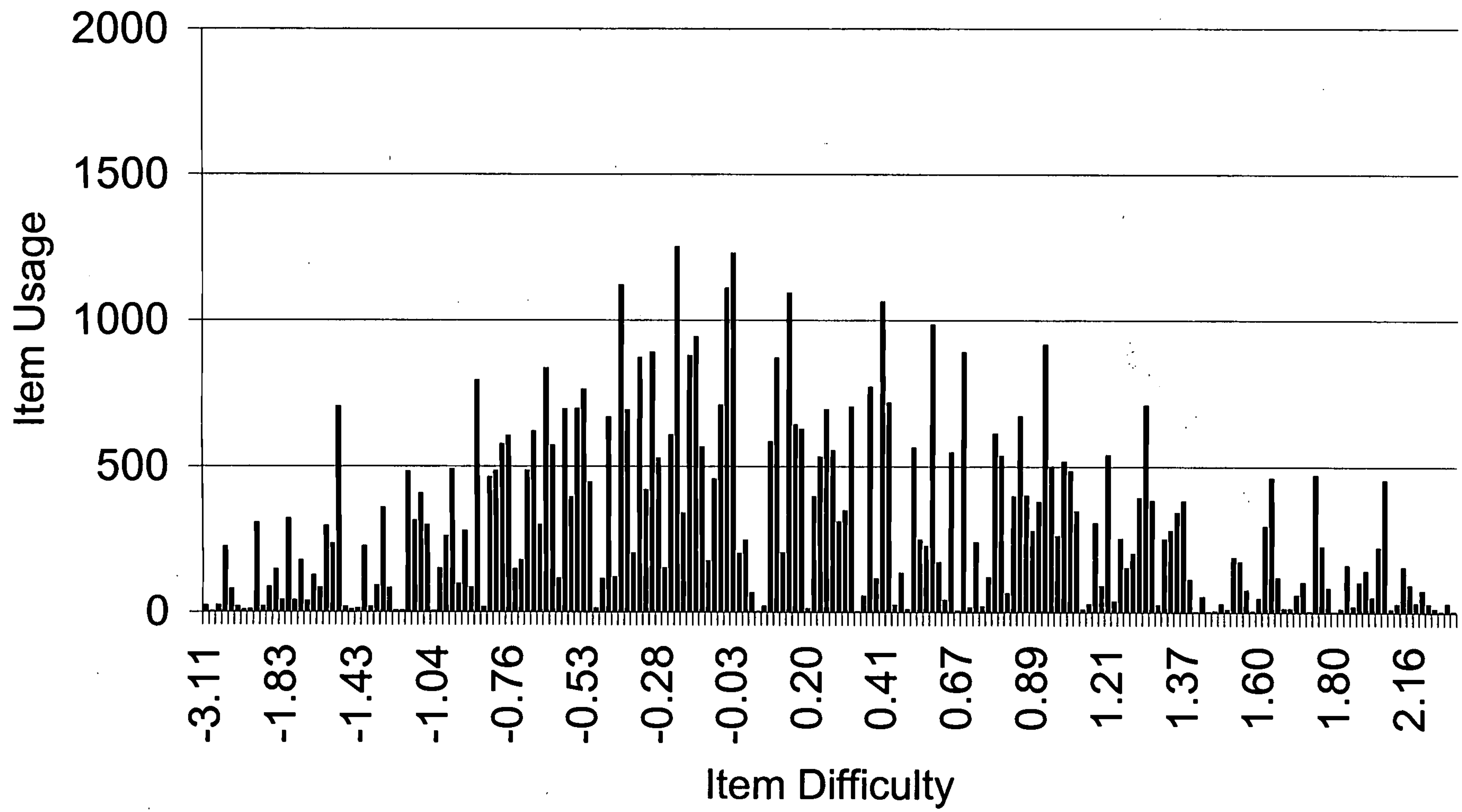


Figure 8. Item Usage Distribution on 30-Item Test (Exposure Control at 0.90)--Cours&GPA Method

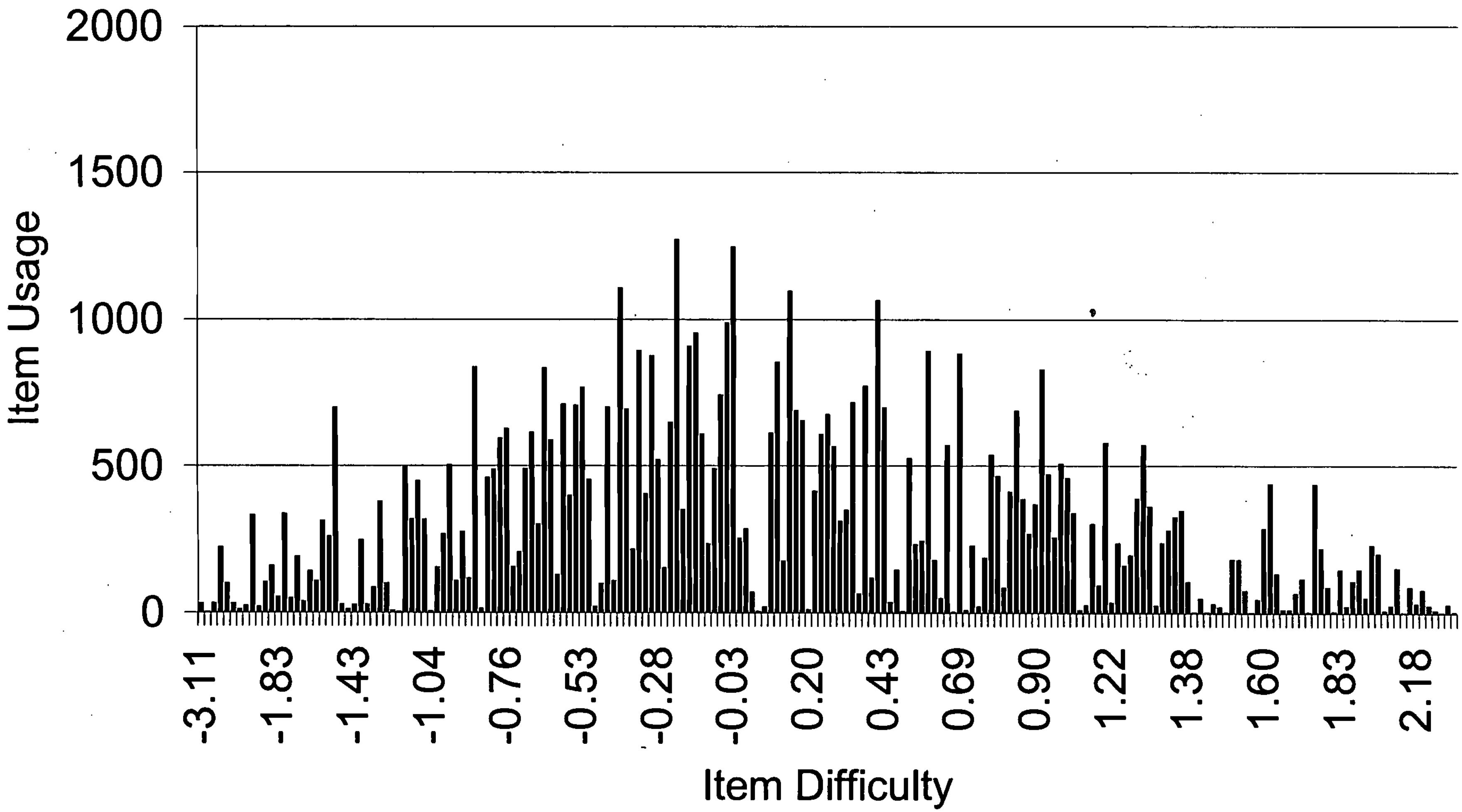


Figure 9. Item Usage Distribution on 15-Item Test (Exposure Control at 0.10)--No-Info Method

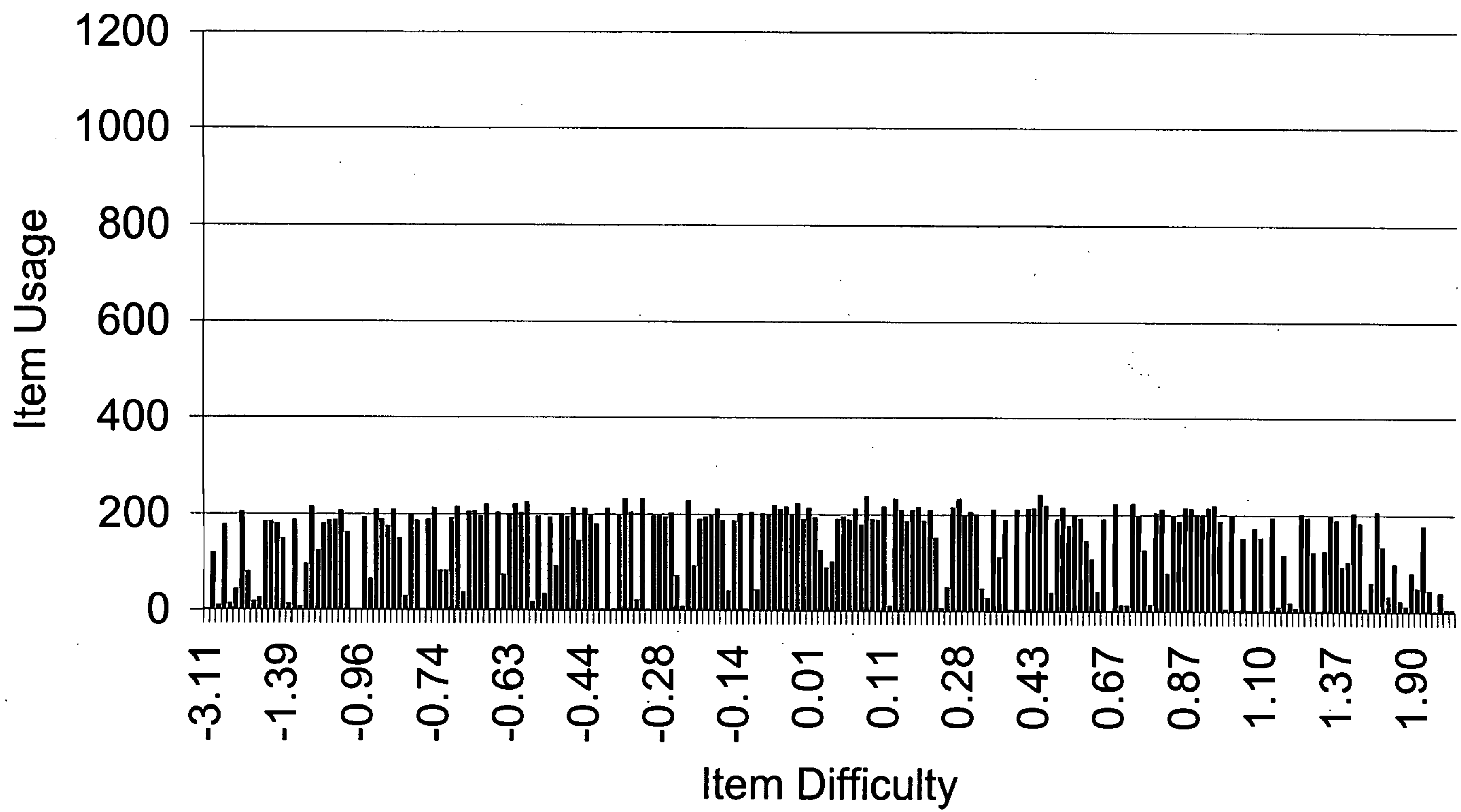


Figure 10. Item Usage Distribution on 15-Item Test
(Exposure Control at 0.10)--GPA Method

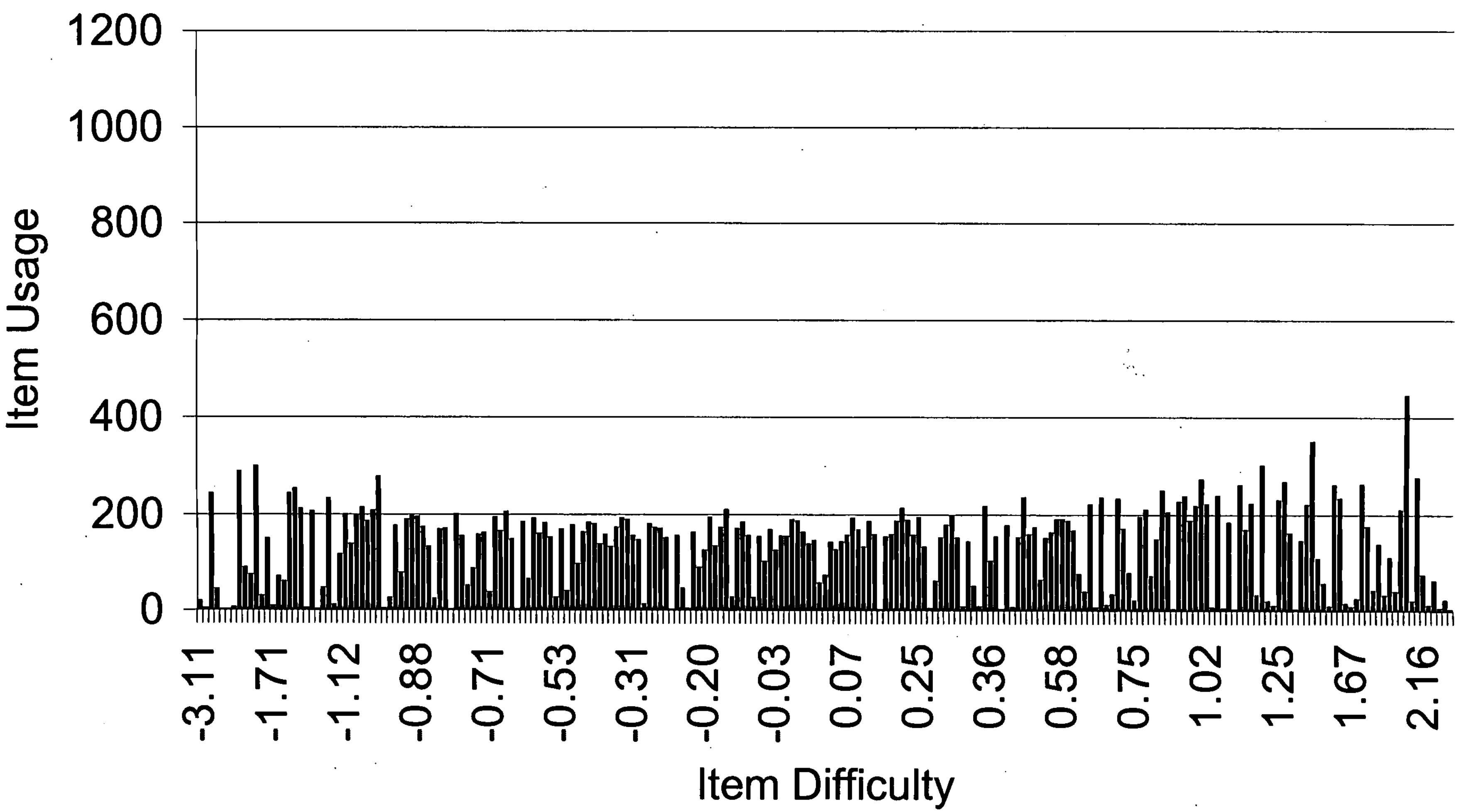


Figure 11. Item Usage Distribution on 15-Item Test (Exposure Control at 0.10)--Course&GPA Method

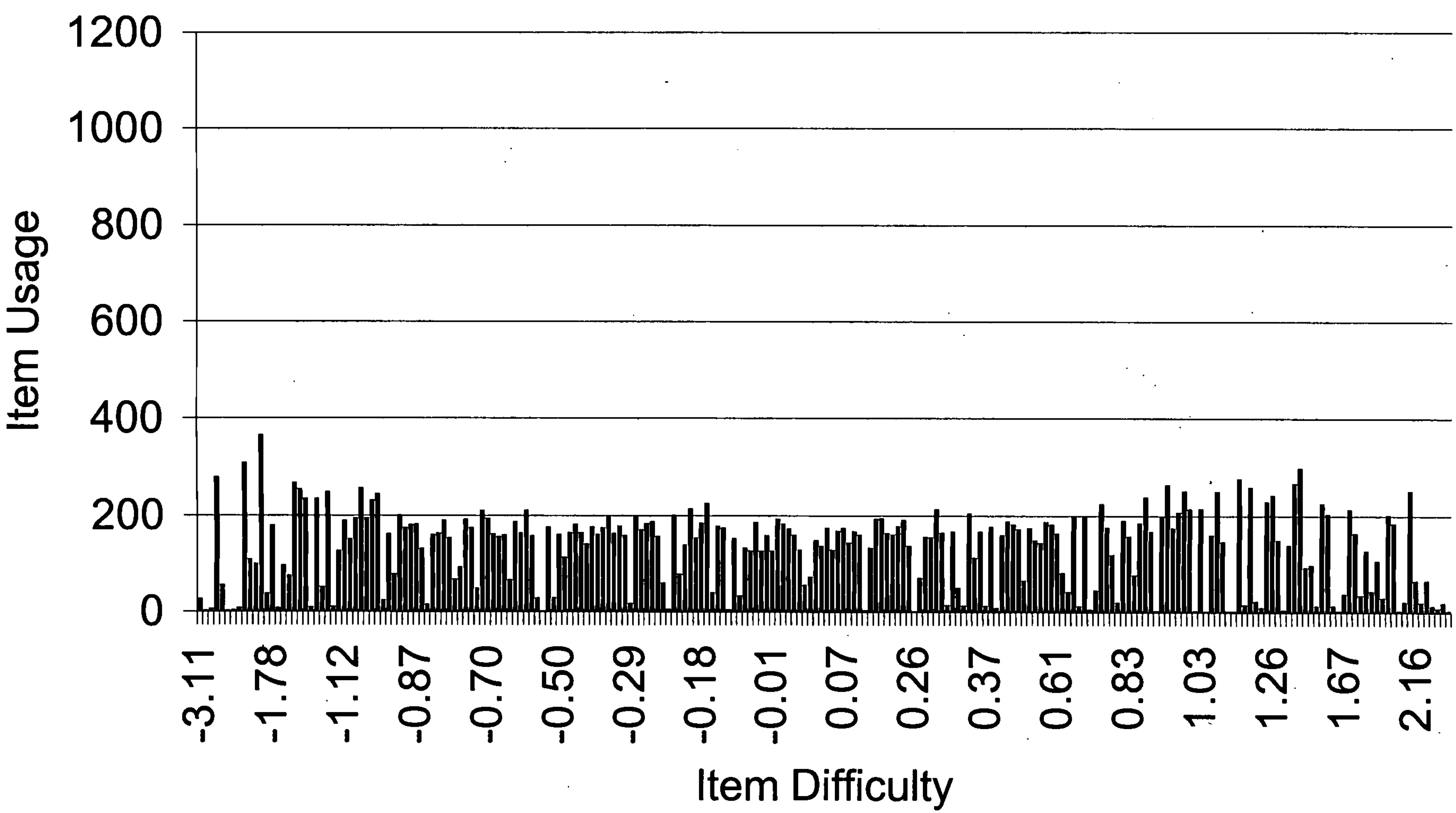


Figure 12. Item Usage Distribution on 30-Item Test (Exposure Control at 0.10)--No-Info Method

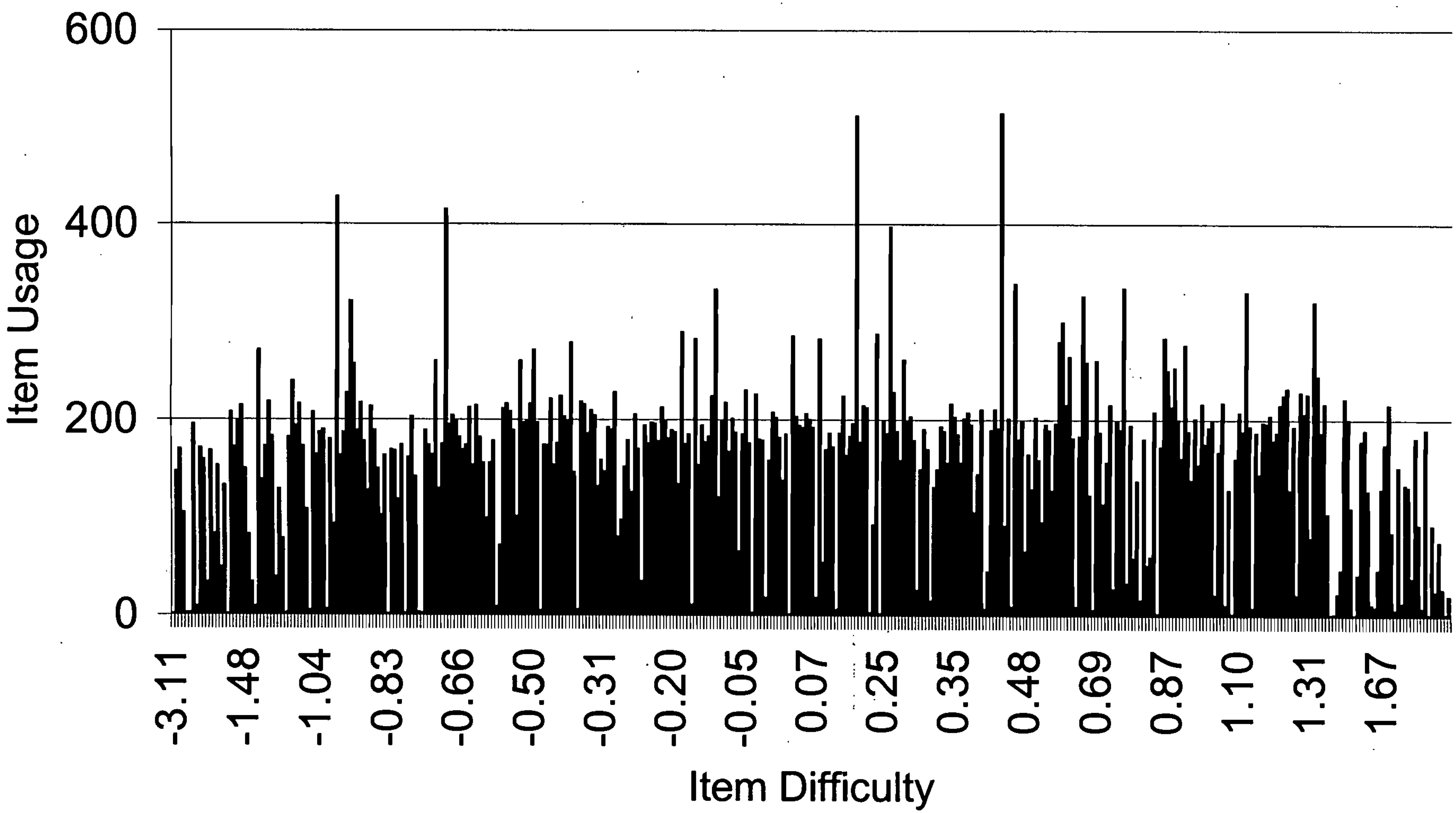


Figure 13. Item Usage Distribution on 30-Item Test (Exposure Control at 0.10)--GPA Method

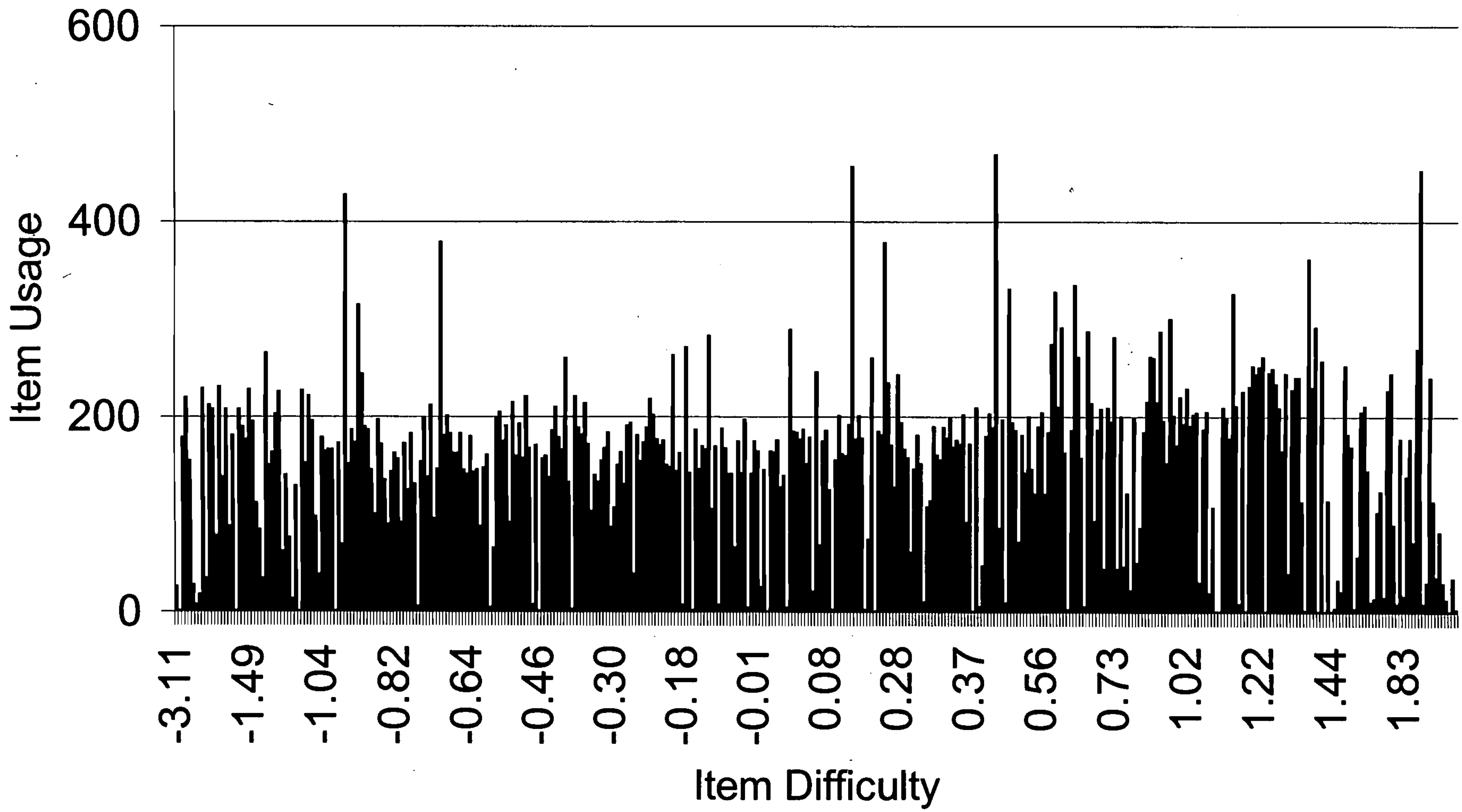


Figure 14. Item Usage Distribution on 30-Item Test (Exposure Control at 0.10)--Course&GPA Method

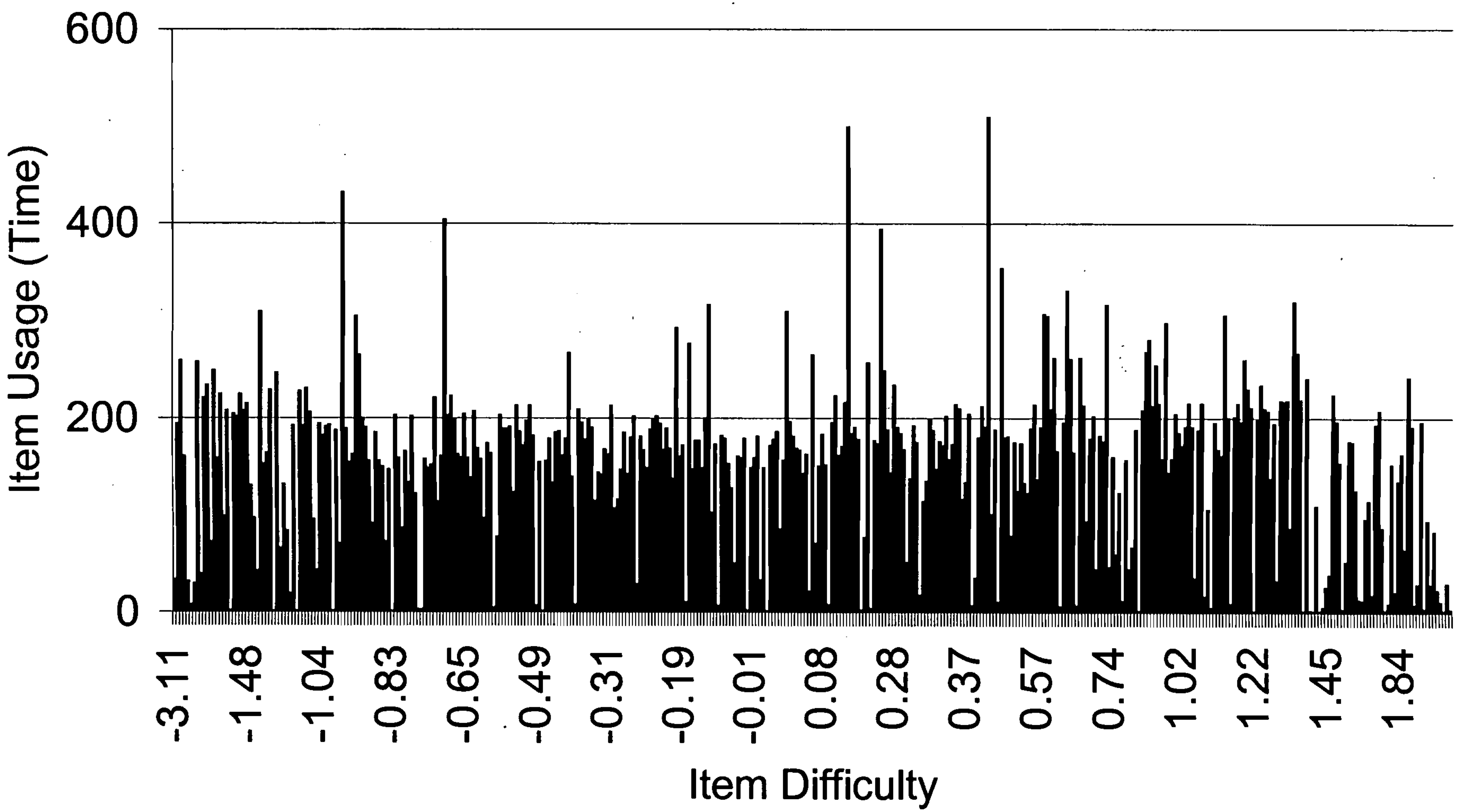


Figure 15. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.90) --No-Info Method & High Precision

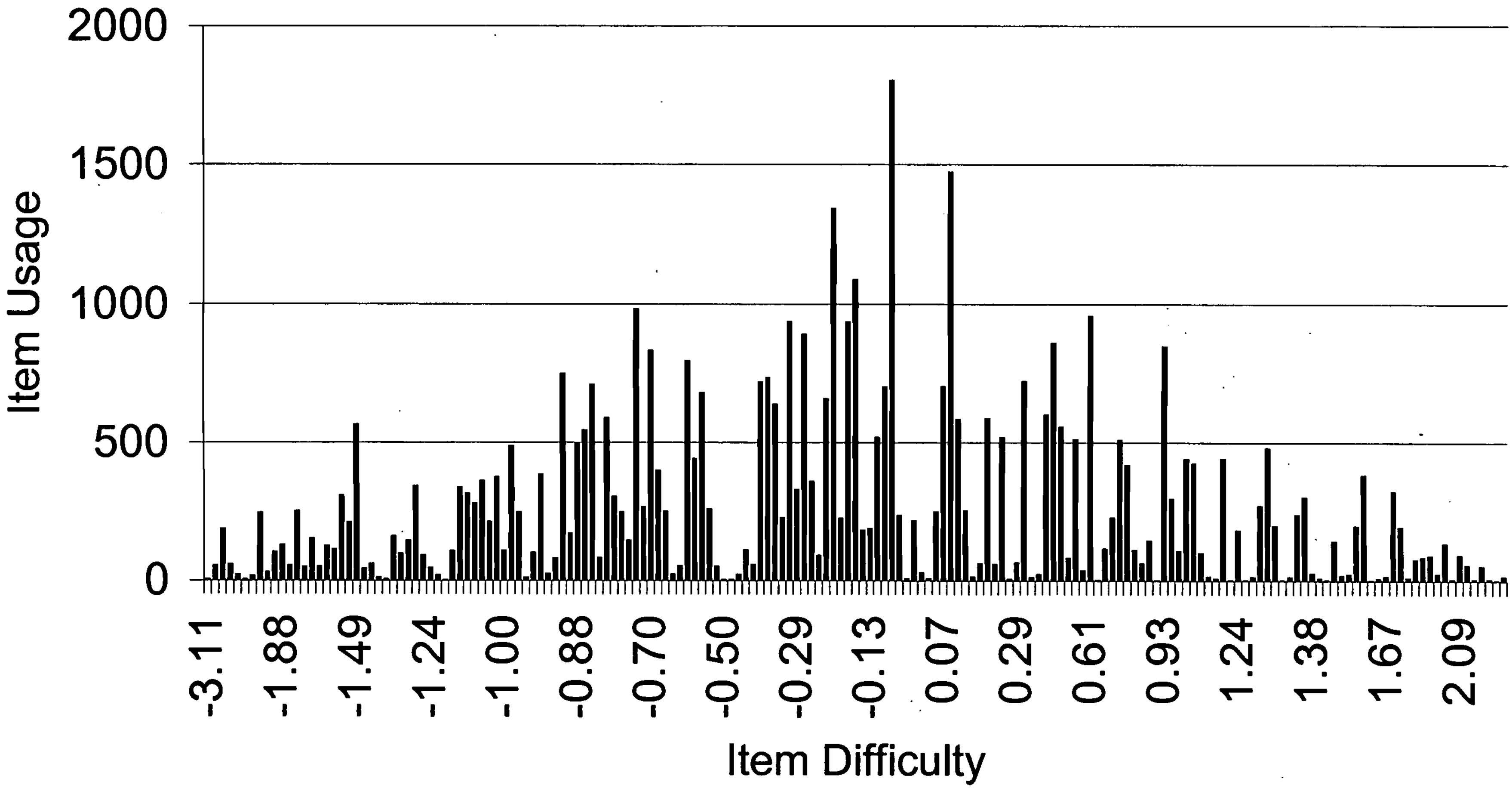


Figure 16. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.90) --GPA Method & High Precision

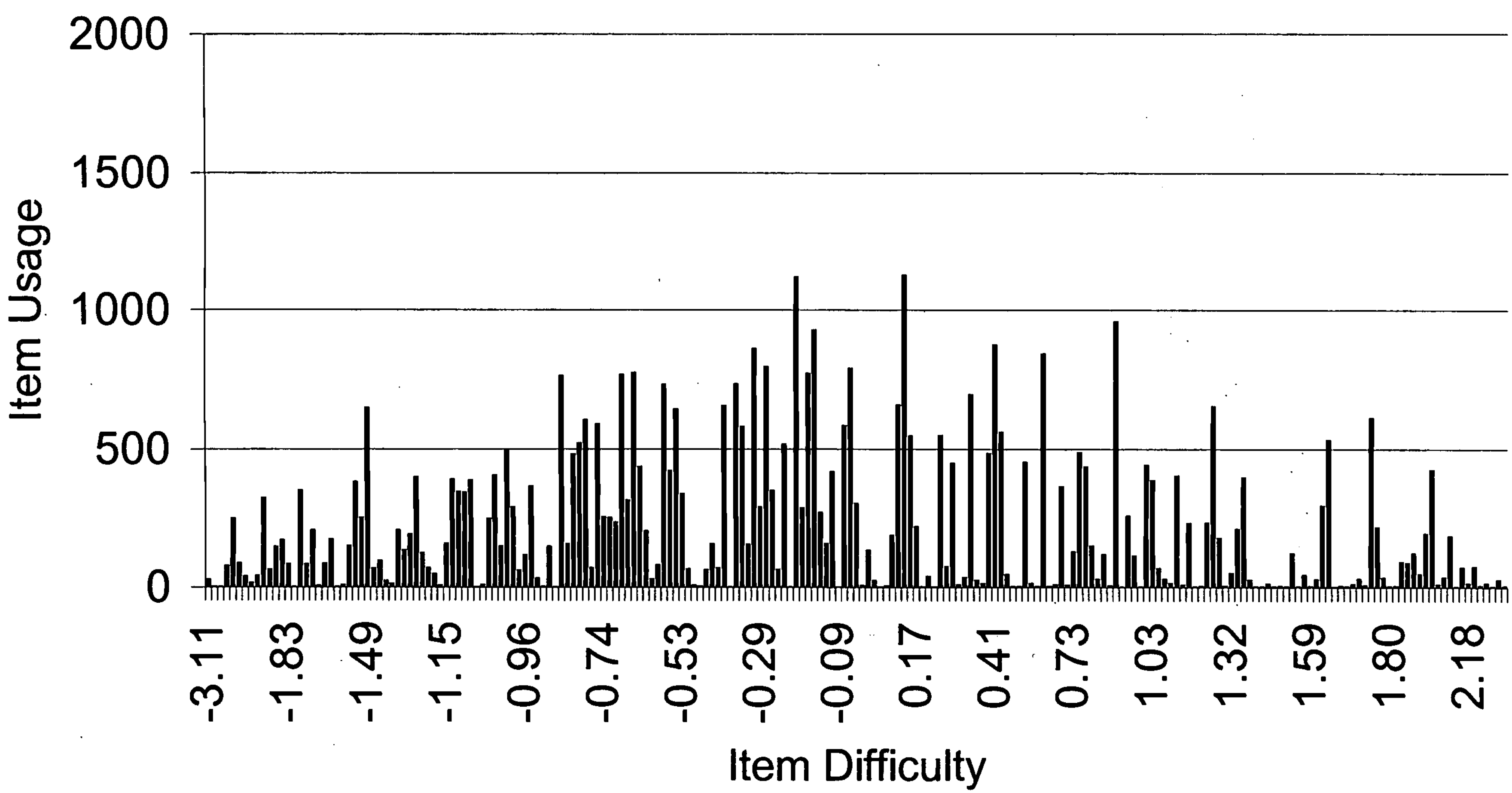
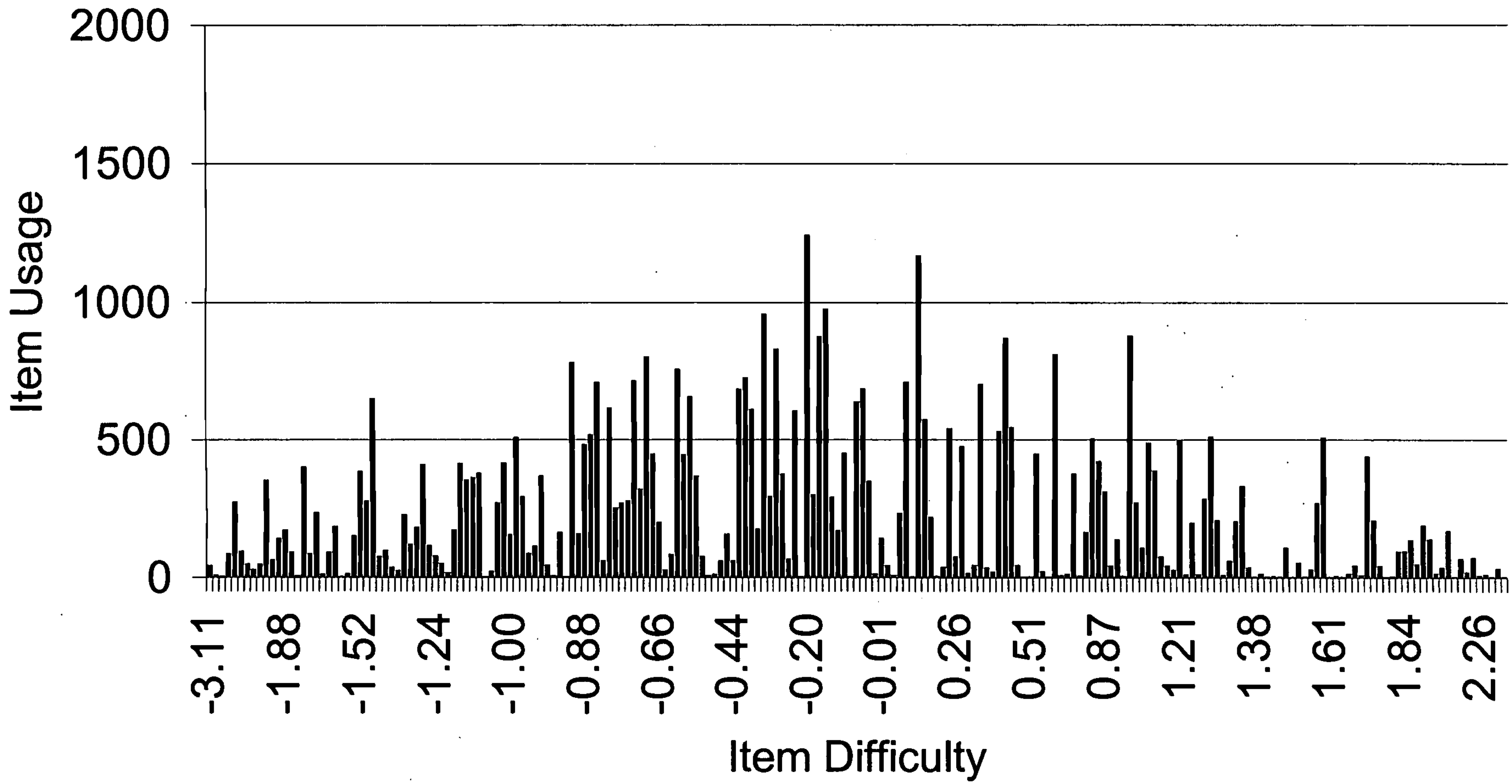


Figure 17. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.90)--Cours&GPA Method & High Precision



BEST COPY AVAILABLE

Figure 18. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.10) --No-Info Method & High Precision

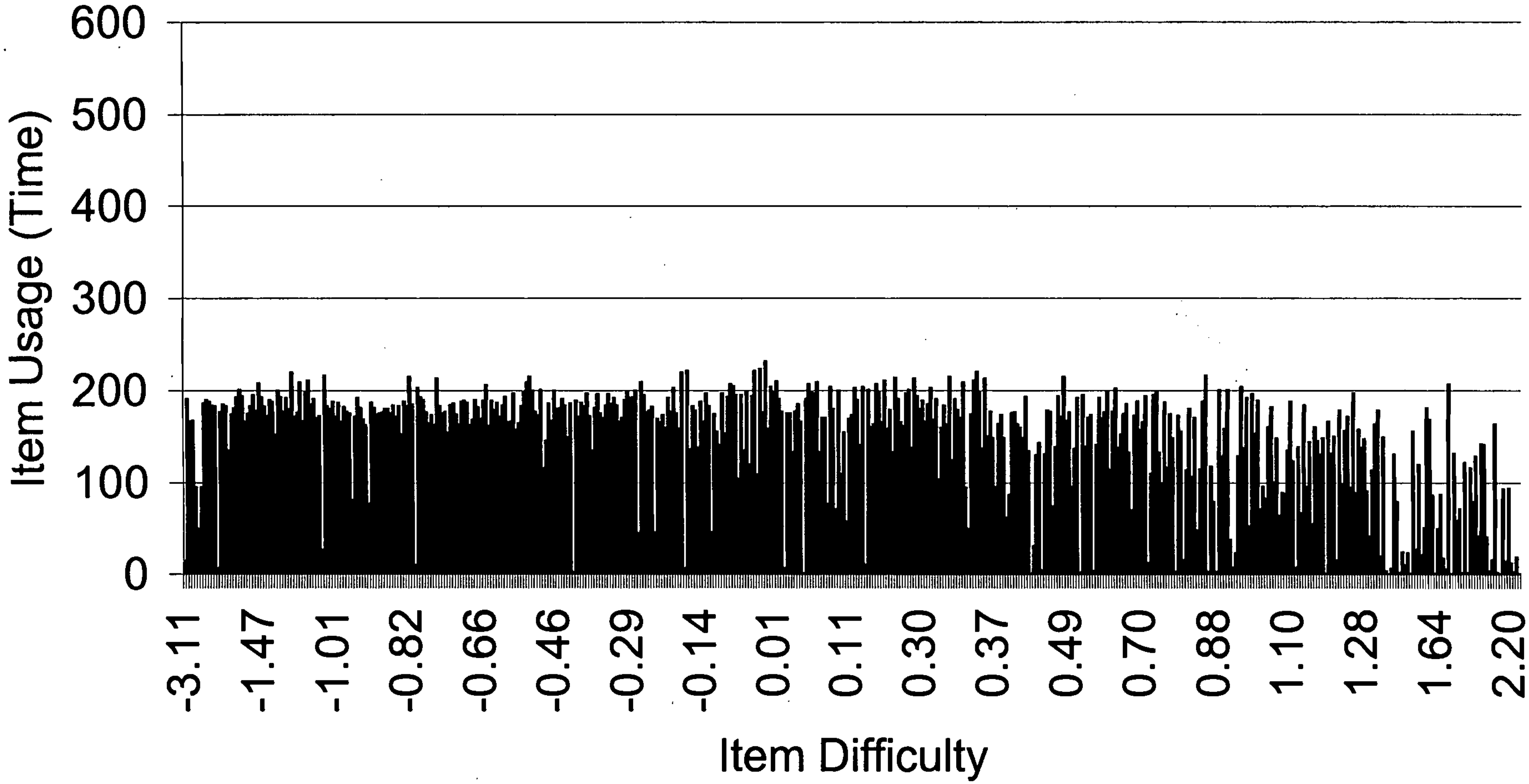


Figure 19. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.10) --GPA Method & High Precision

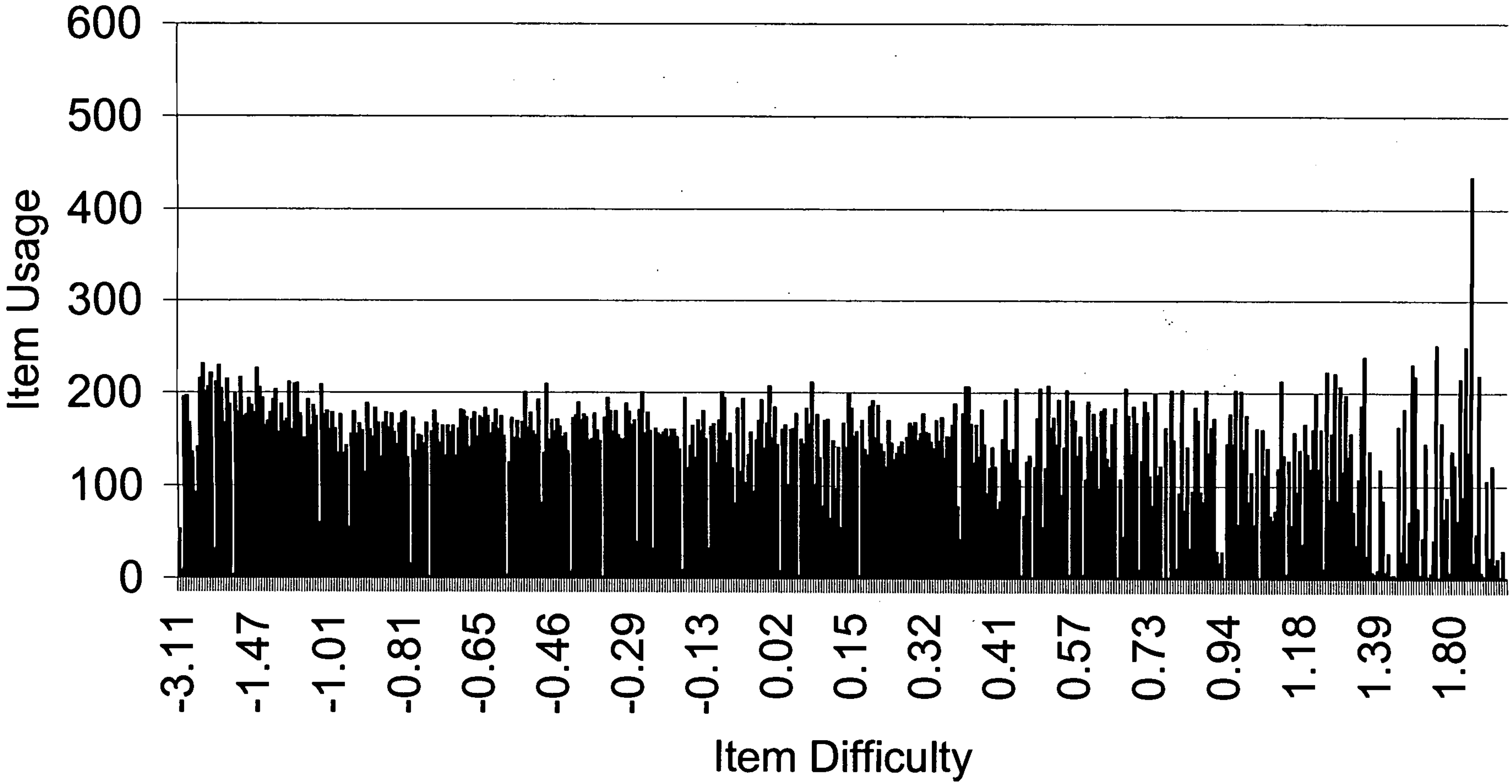
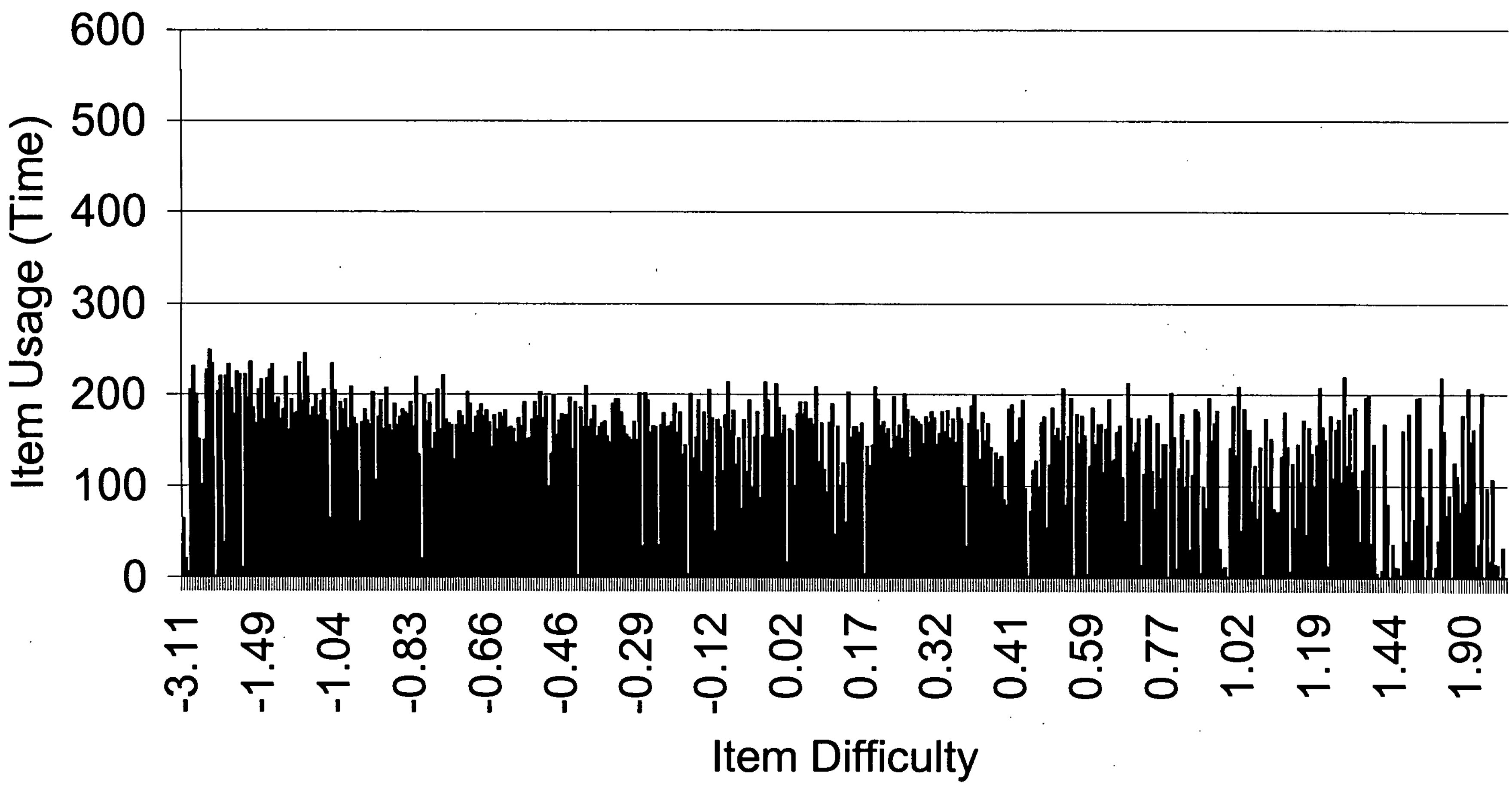


Figure 20. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.10)--Course&GPA Method & High Precision



BEST COPY AVAILABLE

Figure 21. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.90) --No-Info Method & Low Precision

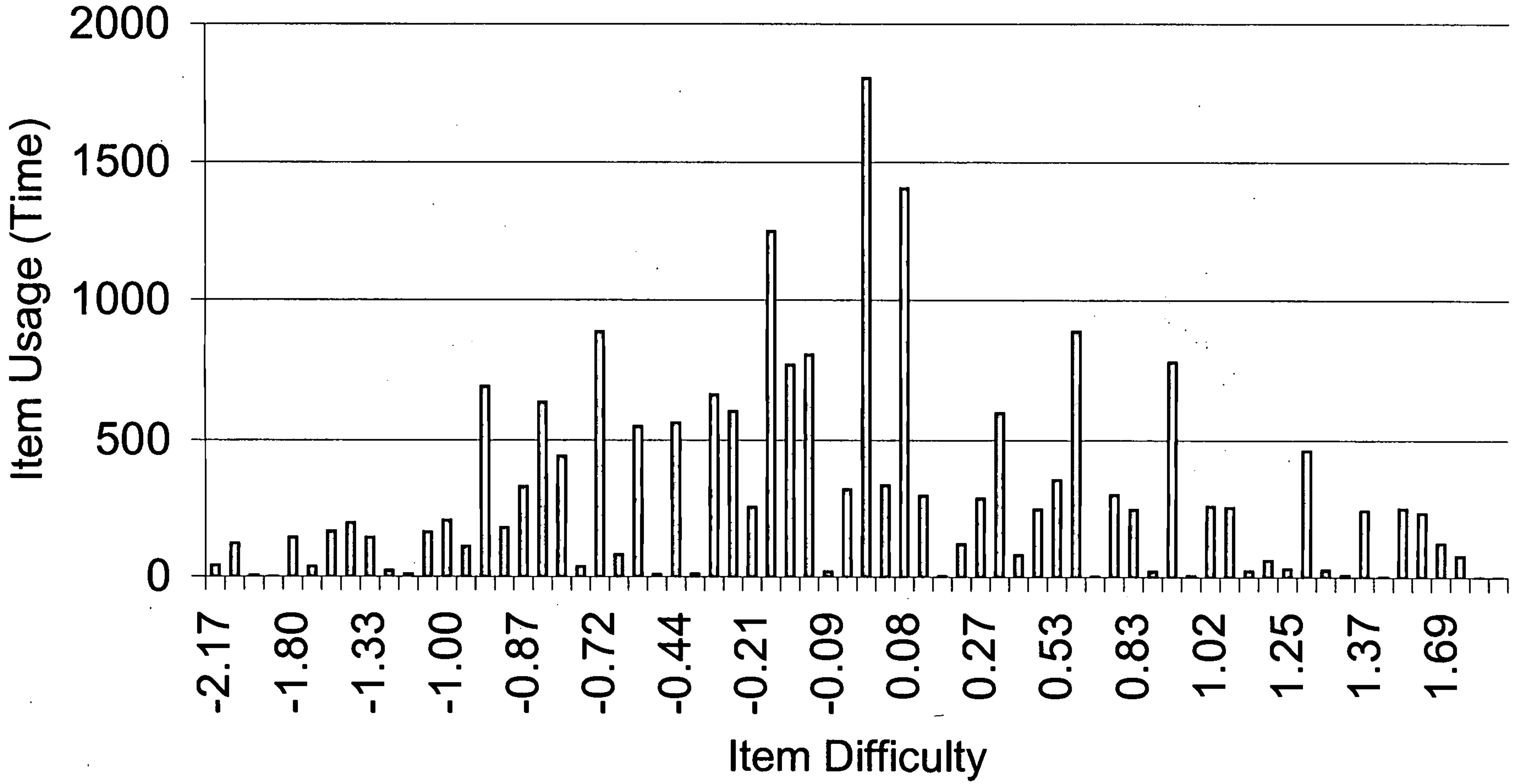


Figure 22. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.90) --GPA Method & Low Precision

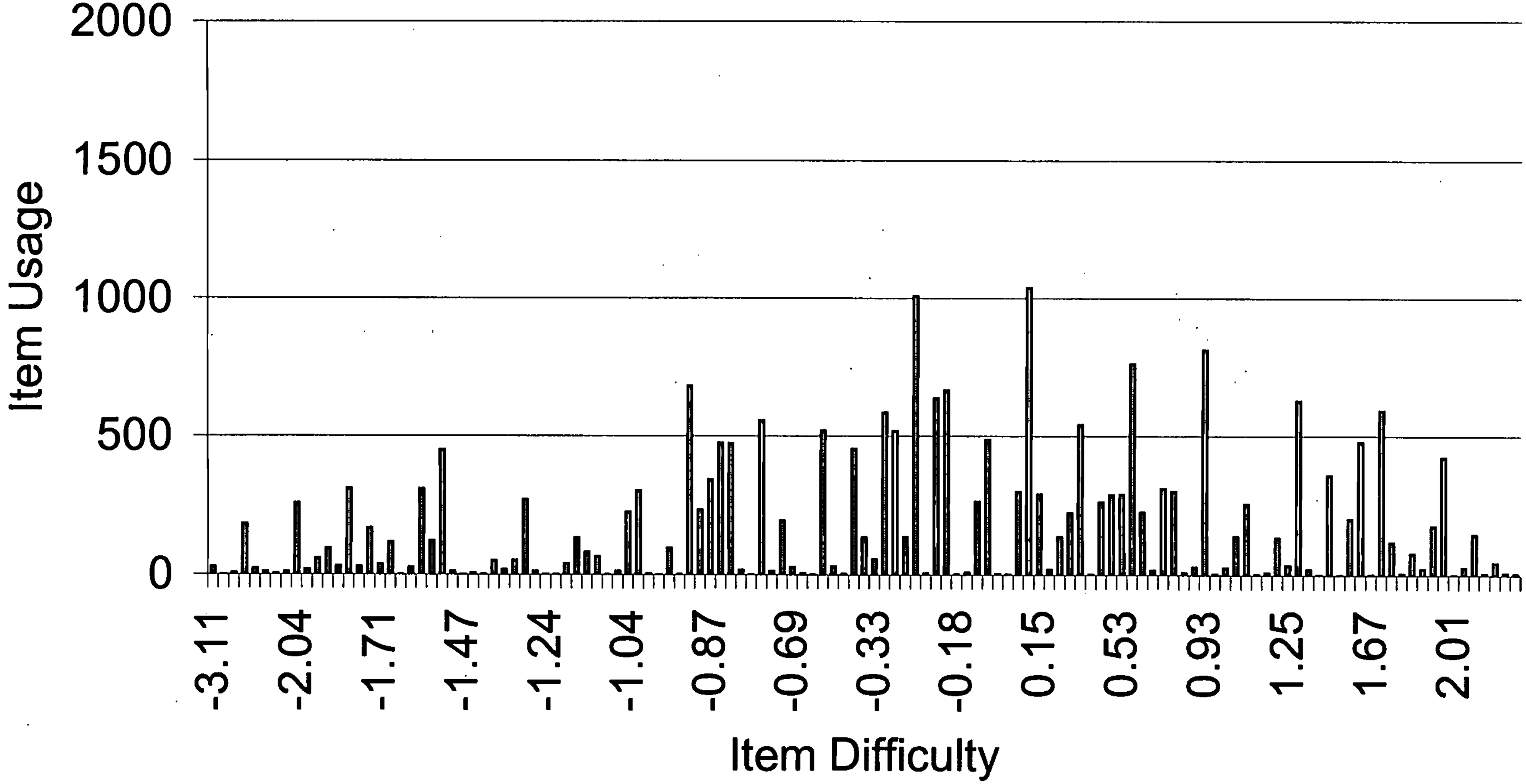
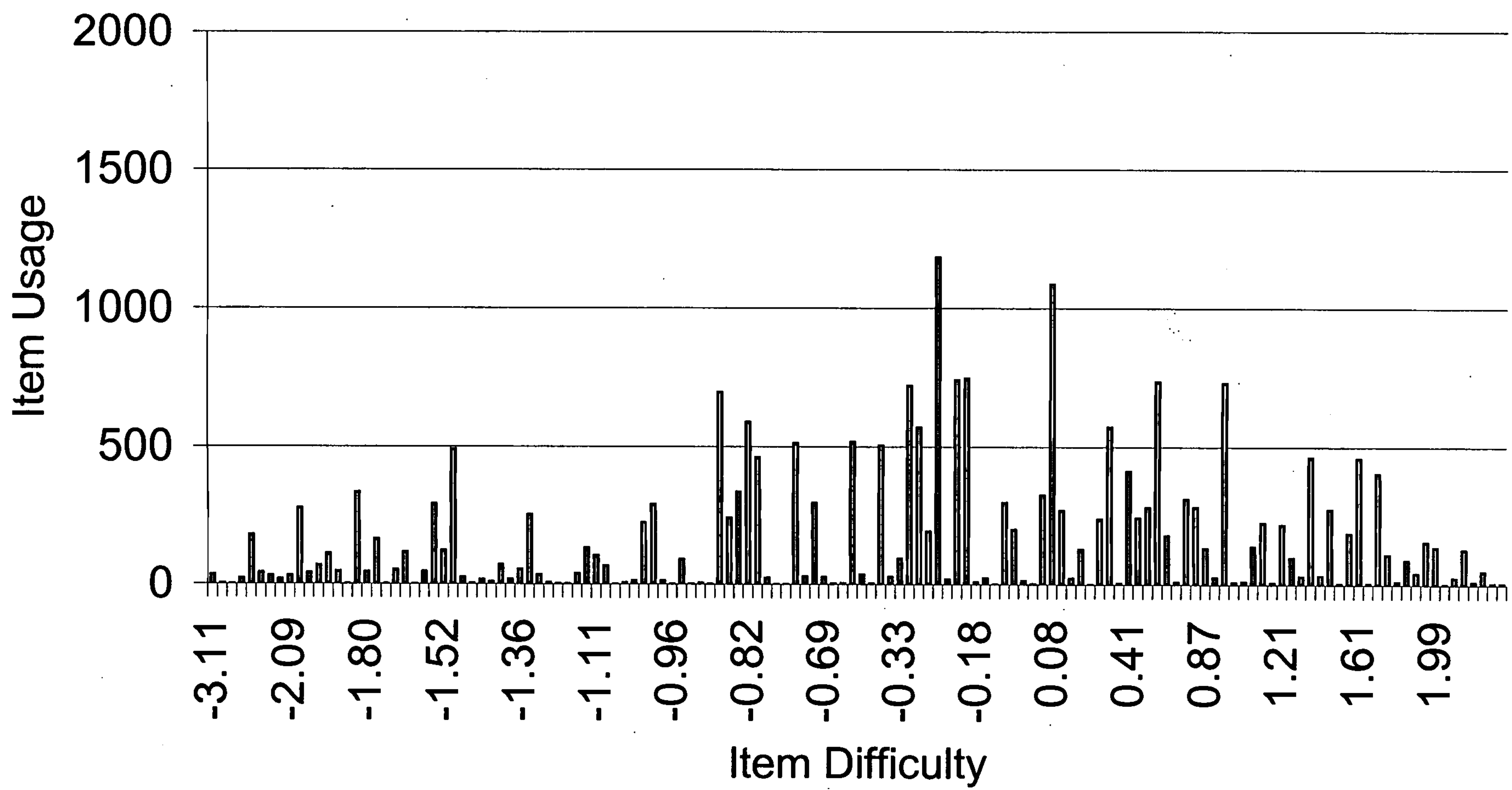


Figure 23. Item Usage Distribution on Variable-Length 45-Item Max Test (Exposure Control at 0.90)--Course&GPA Method & Low Precision



BEST COPY AVAILABLE

Figure 24. Item Usage Distribution on Variable Length 45-Item-Max Test (Exposure Control at 0.10) --No-Info Method & Low Precision

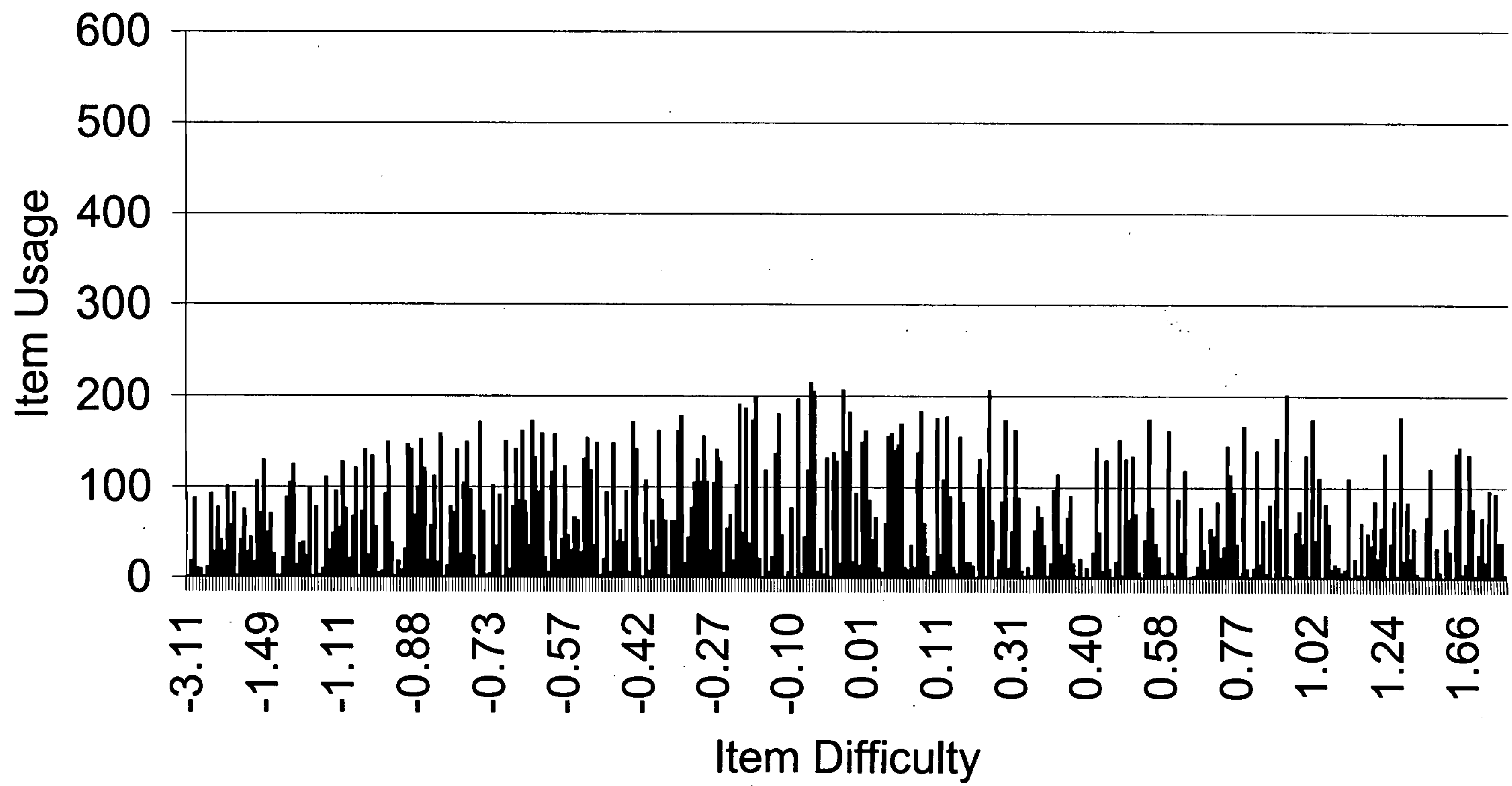


Figure 25. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.10) --GPA Method & Low Precision

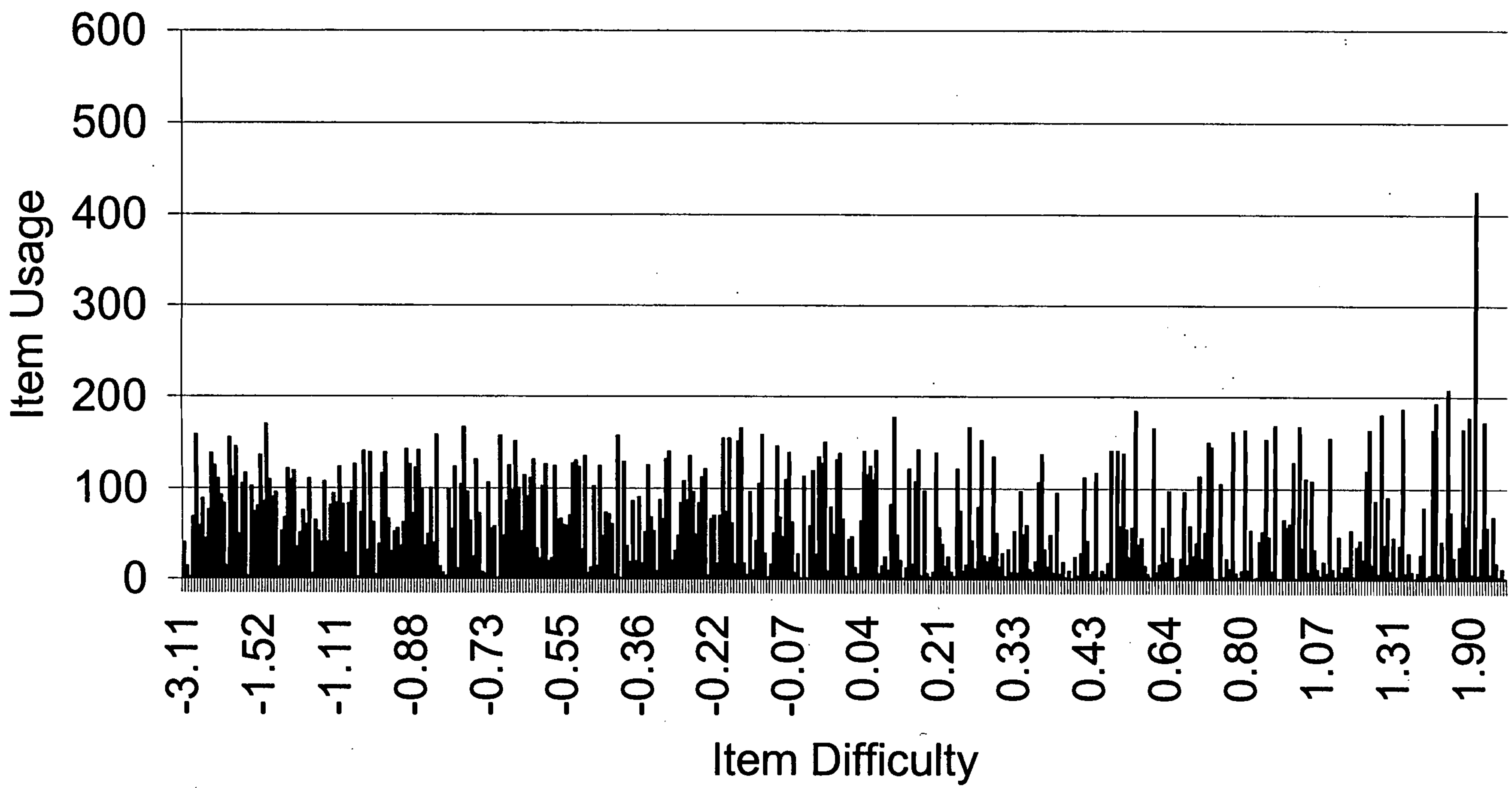
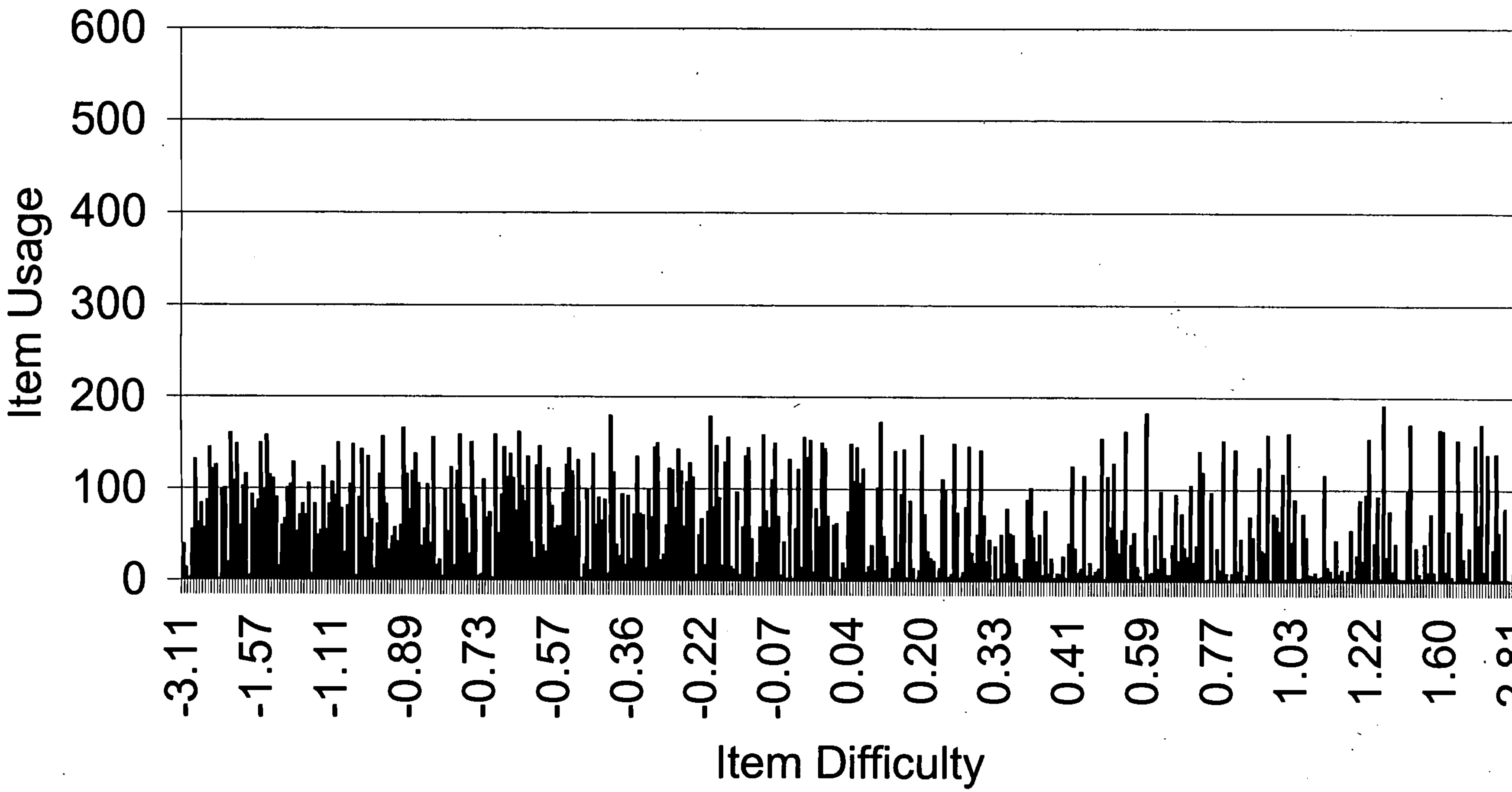


Figure 26. Item Usage Distribution on Variable-Length 45-Item-Max Test (Exposure Control at 0.10)--Course&GPA Method & Low Precision



BEST COPY AVAILABLE

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM029707

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Adjusting Computer Adaptive Test Starting Points to Conserve Item Pool

Author(s): Daming Zhu and Meichu Fan

Corporate Source: ACT, Inc.

Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1	2A	2B
Level 1	Level 2A	Level 2B
☒	☐	☐
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here, → please

Signature:	Printed Name/Position/Title: Daming Zhu	
Organization/Address: ACT, Inc., P.O. Box 168, Iowa City, IA 52243-0168	Telephone: (319) 354-8951	FAX: (319) 339-3021
	E-Mail Address: zhu@act.org	Date: 4/12/1999

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

THE UNIVERSITY OF MARYLAND
ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
1129 SHRIVER LAB, CAMPUS DRIVE
COLLEGE PARK, MD 20742-5701
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
1100 West Street, 2nd Floor
Laurel, Maryland 20707-3598

Telephone: 301-497-4080
Toll Free: 800-799-3742
FAX: 301-953-0263
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com